

# Automated avalanche mapping with deep learning: from satellite to webcam imagery

Elisabeth D. Hafner



Diss. ETH No. 30278

# Automated avalanche mapping with deep learning: from satellite to webcam imagery

*A thesis submitted to attain the degree of*

DOCTOR OF SCIENCES  
(Dr. sc. ETH Zurich)

*presented by*

Elisabeth Doris HAFNER  
MSc. GEOSPATIAL TECHNOLOGIES, TU GRAZ  
born on 26.10.1991

*accepted on the recommendation of*

Prof. Dr. Konrad Schindler  
Prof. Dr. Jan Dirk Wegner  
Dr. Yves Bühler  
Assoc. Prof. Dr. Pascal Haegeli  
Adj. Assoc. Prof. Dr. Regula Frauenfelder

2024



# Abstract

Humans have been exposed to snow avalanches ever since they inhabited or travelled through mountainous regions. For centuries, ways of managing risk and mitigating damage due to avalanches have been limited. For example, forests prevented the release of avalanches, protecting the villages below. In the past ca. 150 years, avalanche protection measures and risk reduction strategies have become increasingly sophisticated. With a growing population, the rise of alpine tourism, and major transportation corridors running through the mountains, dealing with avalanches has become more important than ever. Operational services such as the avalanche warning service, hazard mapping, or the installation of protection and mitigation measures for endangered zones are important tools for avalanche risk management. The efficiency of these tools depends on knowledge about past avalanche occurrences. In this thesis, novel methods are developed to automatically provide this information over large regions. Specifically, state-of-the-art deep learning technology is used to extract such information from optical satellite and webcam imagery. The avalanches required to train the deep learning models are identified and mapped by human experts in the domain. Therefore, the thesis also focuses on the consistency of the avalanche area identified by various domain experts.

In the first part of the thesis, a DeeplabV3+ model is adapted to automatically identify and map avalanches from optical SPOT 6/7 satellite imagery (1.5 m resolution). The model is trained, validated, and tested with more than 24'000 manually annotated avalanche polygons. The data originate from two avalanche periods in January 2018 and January 2019 and cover an area of more than 22'000 km<sup>2</sup>. In addition, the quality of the model and the reproducibility of avalanches manually annotated by experts are assessed for a small subset of the data.

The second part of the thesis investigates in more detail the reproducibility of estimates of avalanche dimensions by human experts in three user studies. The first study analyzes the classification of ten avalanches into five standardized size categories by each of 170 avalanche experts from Europe and North America. The second and the third study examine avalanches manually mapped from oblique photographs (6 avalanches, 10 participants) or from remotely sensed imagery (2.9 km<sup>2</sup>, 5 participants), respectively.

The third part of the thesis leverages interactive avalanche segmentation (IAS) to combine human expert knowledge with deep learning. Here, when mapping avalanches from webcam imagery, the user collaborates with the previously trained model. The use of the model is supposed to make avalanche mapping more accurate and efficient. For this purpose, we adapt a state-of-the-art interactive segmentation model based on HRNet+OCR and train it for avalanche segmentation from webcam imagery. The human user interacts with the model through confirming or corrective feedback.

In summary, this thesis makes a substantial contribution to the development of an operational automatic avalanche mapping service. The thesis provides the first automatic avalanche

mapping from optical satellite imagery with deep learning, it quantifies, for the first time to this extent, the reproducibility of human avalanche estimates, and it presents a first interactive approach for the mapping of avalanches from webcam imagery. Thereby, this thesis contributes to a more efficient use of data and better provision of information on past avalanche occurrences, and thus it benefits decision making in safety-relevant applications.

# Zusammenfassung

Seit der Mensch in Bergregionen lebt oder durch sie reist, hat er mit Lawinen zu tun. Jahrhundertlang waren die Möglichkeiten zur Risikosenkung und zum Schutz vor Lawinen begrenzt. Zum Beispiel verhinderten Wälder den Abgang von Lawinen und schützten dadurch die darunterliegenden Dörfer. Lawinenschutzmassnahmen und weitere Möglichkeiten zur Risikominimierung haben sich in den letzten etwa 150 Jahren stark weiterentwickelt. Der Umgang mit Lawinen ist heute aufgrund der gewachsenen Bevölkerung, des Tourismus und des Baus neuer Verkehrsachsen in Bergregionen mehr denn je von entscheidender Bedeutung. Dabei sind der Lawinenwarndienst, die Gefahrenkartierung oder die Errichtung von Schutzmassnahmen in gefährdeten Gebieten wichtige Instrumente des Risikomanagements. Die Wirksamkeit dieser Instrumente basiert zu einem grossen Anteil auf Informationen über vergangene Lawinenereignisse. In dieser Arbeit werden neue Methoden entwickelt, um solche Informationen automatisch über grössere Regionen bereitzustellen. Konkret werden mit modernen Deep-Learning-Methoden Lawinen aus optischen Satellitenbildern und aus Webcambildern extrahiert. Die für das Training der Deep-Learning-Methoden verwendeten Lawinen werden von menschlichen Experten eingezeichnet. Daher befasst sich ein weiterer Teil der Arbeit mit der Übereinstimmung der von verschiedenen Experten geschätzten Ausdehnung einzelner Lawinen.

Im ersten Teil der Arbeit wird ein DeeplabV3+ Modell zur automatischen Identifizierung und Kartierung von Lawinen aus optischen SPOT 6/7-Satellitenbildern (1.5 m Auflösung) vorgestellt. Das Modell wurde mit über 24'000 manuell kartierten Lawinenpolygonen trainiert, validiert und getestet. Diese Kartierungen stammen von zwei Lawinenperioden im Januar 2018 und Januar 2019 und decken eine Fläche von über 22'000 km<sup>2</sup> ab. Zusätzlich wurden die Qualität des Modells und die Reproduzierbarkeit der manuellen Erfassung von Lawinengrenzen durch Experten in einem Teilgebiet untersucht.

Der zweite Teil der Arbeit geht mit drei Studien noch näher auf die Reproduzierbarkeit der Einschätzung von Lawineneigenschaften durch verschiedene Experten ein. Die erste Studie analysiert die Einteilung von zehn Lawinen durch je 170 Lawinenexperten aus Europa und Nordamerika in fünf standardisierte Grössenkatgorien. Die zweite und dritte Studie analysieren manuell auf Karten übertragene Lawinenumrisse aus Schrägfotos (6 Lawinen, 10 TeilnehmerInnen) bzw. aus Fernerkundungsdaten (2,9 km<sup>2</sup>, 5 TeilnehmerInnen).

Der dritte Teil der Arbeit kombiniert mittels interaktiver Lawinensegmentierung menschliches Expertenwissen mit Deep-Learning. Dabei arbeitet der Nutzer beim Einzeichnen von Lawinenumrissen in Webcambildern mit einem zuvor trainierten Modell zusammen. Das Modell soll dabei den zeitlichen Aufwand für die genaue Erfassung von Lawinen reduzieren. Dafür passen wir ein bestehendes Modell auf der Basis von HRNet+OCR an und trainieren es speziell für Lawinen auf Webcambildern. Der Nutzer interagiert dann mit bestätigenden oder korrigierenden Rückmeldungen mit dem Modell.

Insgesamt leistet diese Arbeit einen wichtigen Beitrag zur Entwicklung einer automatisierten operationellen Lawinenkartierung. Sie stellt die erste vollautomatisierte Lawinenkartierung aus optischen Satellitenbildern mit Deep learning bereit, sie quantifiziert zum ersten Mal im vorgelegten Umfang die Reproduzierbarkeit der von menschlichen Experten gemachten Lawineneinschätzung, und sie präsentiert einen ersten interaktiver Ansatz zur Kartierung von Lawinen aus Webcambildern. Damit trägt die Arbeit zu einer effizienteren Aufbereitung und Bereitstellung von Informationen über Lawinenereignisse bei und verbessert dadurch die Grundlage zur fundierten Entscheidungsfindung in sicherheitsrelevanten Anwendungen.



# Acknowledgements

I am very grateful to everyone who supported me in the pursuit of this thesis in the last years: I especially wish to thank my supervisor at SLF, Yves Bühler, for making this work possible and for continuous support. I am grateful to my supervisors Konrad Schindler (ETH Zurich) and Jan Dirk Wegner (University of Zurich and ETH Zurich) for taking me on as an external PhD student as well as for productive exchange, helpful advice, and a refreshing perspective from outside the “avalanche research bubble”.

Special thanks goes to Thomas Stucki and the whole avalanche warning team for igniting my passion for mapping and documenting avalanches in the first place. I am especially grateful to Frank Techel for fruitful discussions and valuable perspectives from the avalanche warning service. Thank you to my team AlpRS for your input, feedback, support, for the great field days and the team excursions. I especially thank Gwendolyn Dasser for productive scientific exchange as well as sharing sarcasm and humor during some of the most stressful times!

I wish to thank all my friends and colleagues at SLF for beneficial exchange, for adding to the big picture, for great support, and wonderful times together. I am especially grateful to the colleagues (in and outside SLF) taking the time for participating in my various user studies and the survey. I am grateful to Konrad’s and Jan’s groups, especially Rodrigo and Dora, for making me feel home at ETH Zurich and for insightful discussions and collaborations in the field of machine learning, as well as for great retreats.

I am grateful to everyone who read a pre-final version of my thesis, helped to spot typos, and made suggestions for improvements.

I am thankful to my parents for fostering a curious mind that aims to explore and understand things. I am grateful to them and to the rest of my family for the continuous support on my journey towards a PhD. The thesis would not have been possible without times of recharging the batteries outside of work: I would like to especially thank Isabelle, Thomi, Simon, my family and all other friends for the small and big adventures in the mountains on skis, on foot, or with Osk, as well as community indoors. Finally, I would like to thank Simon for his unconditional love, for empathy, for encouragement, for support, and for intellectual camaraderie.

Merci eu allne vielvielmol!

# Contents

<b>Abstract</b>	<b>i</b>
<b>Zusammenfassung</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>List of Acronyms</b>	<b>x</b>
<b>I Introduction</b>	<b>1</b>
<b>1 Living with avalanches</b>	<b>2</b>
1.1 Research questions . . . . .	3
<b>2 Classification and definition of avalanches</b>	<b>4</b>
2.1 Avalanche types . . . . .	4
2.1.1 Release type . . . . .	4
2.1.2 Type of trigger . . . . .	4
2.1.3 Type of movement . . . . .	5
2.1.4 Liquid water content . . . . .	6
2.2 Avalanche parts . . . . .	6
2.3 Avalanche size . . . . .	6
<b>3 Destructive avalanche periods in the past</b>	<b>8</b>
<b>4 Applications depending on proper avalanche information</b>	<b>10</b>
4.1 Avalanche warning service . . . . .	10
4.2 Avalanche modeling . . . . .	12
4.3 Hazard mapping . . . . .	13
4.4 Planning and evaluation of protection/mitigation measures . . . . .	14
4.5 Risk analysis and safety concepts . . . . .	15
<b>5 Documentation of avalanches</b>	<b>17</b>
5.1 Historical events and data from the past . . . . .	17
5.2 Avalanche documentation and mapping today . . . . .	17
5.2.1 Observers . . . . .	18
5.2.2 Seismic sensors . . . . .	18
5.2.3 Infrasound . . . . .	18
5.2.4 Doppler radar . . . . .	18

5.2.5	Webcam data . . . . .	18
5.2.6	Satellite data . . . . .	19
5.2.7	Airplane and Drone data . . . . .	20
<b>6</b>	<b>Data and processing methods</b>	<b>22</b>
6.1	Sensors and datasets . . . . .	22
6.1.1	Sentinel-1 . . . . .	22
6.1.2	Sentinel-2 . . . . .	22
6.1.3	SPOT 6/7 . . . . .	23
6.1.4	Ultracam airplane imagery . . . . .	23
6.1.5	Wingtra drone imagery . . . . .	23
6.1.6	Swiss national DEM . . . . .	23
6.1.7	SWISSIMAGE . . . . .	23
6.1.8	Swiss national topographic map . . . . .	24
6.2	Semantic segmentation with neural networks . . . . .	24
6.3	Monophotogrammetry . . . . .	25
<b>7</b>	<b>Preceding work on optical satellite avalanche mapping</b>	<b>26</b>
7.1	Large-scale mapping in Switzerland . . . . .	26
7.2	Investigation into performance and completeness . . . . .	26
<b>II</b>	<b>Automated avalanche mapping from SPOT 6/7 satellite imagery with deep learning: results, evaluation, potential and limitations</b>	<b>28</b>
<b>8</b>	<b>Introduction</b>	<b>30</b>
<b>9</b>	<b>Data</b>	<b>33</b>
<b>10</b>	<b>Method</b>	<b>35</b>
10.1	Model architecture . . . . .	35
10.1.1	Sampling and data split . . . . .	36
10.1.2	Training . . . . .	38
<b>11</b>	<b>Results and discussion</b>	<b>40</b>
11.1	Results and generalization ability . . . . .	40
11.2	Ablation studies . . . . .	43
11.3	Reproducibility of manually mapped avalanches . . . . .	45
11.4	Limitations of this study . . . . .	46
<b>12</b>	<b>Conclusion and outlook</b>	<b>48</b>
<b>III</b>	<b>Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice</b>	<b>49</b>
<b>13</b>	<b>Introduction</b>	<b>51</b>
<b>14</b>	<b>Background</b>	<b>53</b>
<b>15</b>	<b>Data and methods</b>	<b>56</b>

15.1 Study 1: avalanche size estimation . . . . .	56
15.2 Study 2: avalanche mapping from oblique photographs . . . . .	57
15.3 Study 3: avalanche mapping from remotely sensed imagery . . . . .	57
15.4 Data analysis . . . . .	58
15.4.1 Avalanche size estimates . . . . .	58
15.4.2 Avalanche outline determination . . . . .	59
<b>16 Results</b>	<b>60</b>
16.1 Avalanche size estimation (study 1) . . . . .	60
16.2 Avalanche mapping from oblique photographs (study 2) . . . . .	67
16.3 Avalanche mapping from remotely sensed imagery (study 3) . . . . .	69
<b>17 Discussion</b>	<b>73</b>
17.1 Avalanche size estimation (study 1) . . . . .	73
17.2 Avalanche mapping from oblique photographs (study 2) and from remotely sensed imagery (study 3) . . . . .	75
17.3 Implications for practice . . . . .	76
17.4 Limitations . . . . .	77
<b>18 Conclusions and Outlook</b>	<b>79</b>
<b>IV Interactive Snow Avalanche Segmentation from Webcam Imagery: results, potential and limitations</b>	<b>81</b>
<b>19 Introduction</b>	<b>83</b>
<b>20 Data</b>	<b>86</b>
20.1 SLF Webcam network . . . . .	86
20.2 Avalanche images and annotations . . . . .	88
20.2.1 SLF dataset . . . . .	88
20.2.2 UIBK dataset . . . . .	89
<b>21 Methodology</b>	<b>90</b>
21.1 Model architecture . . . . .	90
21.2 Evaluation metrics . . . . .	91
21.2.1 Pixel-wise metrics . . . . .	92
21.2.2 Object-wise metrics . . . . .	92
21.2.3 Comparison of time needed . . . . .	92
21.3 Experimental setup . . . . .	93
21.4 User Study . . . . .	93
<b>22 Results</b>	<b>94</b>
22.1 Pixel-wise metrics . . . . .	94
22.2 Object-wise metrics . . . . .	98
22.3 User study and time saved . . . . .	98
<b>23 Discussion</b>	<b>101</b>
<b>24 Conclusions and Outlook</b>	<b>104</b>

<b>V</b>	<b>Conclusions</b>	<b>106</b>
<b>25</b>	<b>Contributions of this thesis</b>	<b>107</b>
	25.1 Benefits to practice, science and relevance to society . . . . .	108
	25.2 Limitations . . . . .	108
<b>26</b>	<b>Avalanche mapping in the future</b>	<b>110</b>
	26.1 Avalanches in a changing climate . . . . .	110
	26.2 Deep learning aided avalanche mapping . . . . .	110
	26.3 Anticipated advancements of sensors and techniques . . . . .	111
<b>27</b>	<b>Final remarks</b>	<b>113</b>
<b>VI</b>	<b>Backmatter</b>	<b>114</b>
	<b>Appendices</b>	<b>115</b>
<b>A</b>	<b>Appendix to section 6.2</b>	<b>115</b>
<b>B</b>	<b>Appendix to part III</b>	<b>117</b>
<b>C</b>	<b>Supplement to part III</b>	<b>118</b>
	<b>List of Publications</b>	<b>121</b>
	<b>Bibliography</b>	<b>122</b>

# List of Acronyms

<b>AoO</b>	<b>A</b> rea of <b>O</b> verlap
<b>AoU</b>	<b>A</b> rea of <b>U</b> nion
<b>ASPP</b>	<b>A</b> trous <b>S</b> patial <b>P</b> yramid <b>P</b> ooling
<b>BCE</b>	<b>B</b> inary <b>C</b> ross- <b>E</b> ntropy
<b>CNN</b>	<b>C</b> onvolutional <b>N</b> eural <b>N</b> etwork
<b>DAvalMap</b>	<b>DA</b> Vos <b>A</b> valanche <b>M</b> apping Project
<b>DEM</b>	<b>D</b> igital <b>E</b> levation <b>M</b> odel
<b>DSM</b>	<b>D</b> igital <b>S</b> urface <b>M</b> odel
<b>DSPF</b>	<b>D</b> eformable <b>S</b> patial <b>P</b> yramid <b>F</b> low
<b>EAWS</b>	<b>E</b> uropean <b>A</b> valanche <b>W</b> arning <b>S</b> ervices
<b>ESA</b>	<b>E</b> uropean <b>S</b> pace <b>A</b> gency
<b>FOEN</b>	<b>F</b> ederal <b>O</b> ffice for <b>E</b> nvironment
<b>GIS</b>	<b>G</b> eographic <b>I</b> nformation <b>S</b> ystem
<b>GSD</b>	<b>G</b> round <b>S</b> ampling <b>D</b> istance
<b>IAS</b>	<b>I</b> nteractive <b>A</b> valanche <b>S</b> egmentation
<b>IOS</b>	<b>I</b> nteractive <b>O</b> bject <b>S</b> egmentation
<b>IoU</b>	<b>I</b> ntersection <b>o</b> ver <b>U</b> nion
<b>NIR</b>	<b>N</b> ear- <b>I</b> nfra <b>R</b> ed
<b>NoC</b>	<b>N</b> umber <b>o</b> f <b>C</b> licks
<b>OGD</b>	<b>O</b> pen <b>G</b> overnment <b>D</b> ata
<b>POD</b>	<b>P</b> robability <b>o</b> f <b>D</b> etection
<b>PPV</b>	<b>P</b> ositive <b>P</b> redictive <b>V</b> alue
<b>RGB</b>	<b>R</b> ed <b>G</b> reen <b>B</b> lue
<b>RMSE</b>	<b>R</b> oot <b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>SAR</b>	<b>S</b> ynthetic <b>A</b> perture <b>R</b> adar
<b>SLF</b>	<b>W</b> SL Institute for <b>S</b> now and <b>A</b> valanche <b>R</b> esearch <b>SLF</b>
<b>SPOT</b>	<b>S</b> atellite <b>P</b> our l' <b>O</b> bservation de la <b>T</b> erre
<b>SRTM</b>	<b>S</b> huttle <b>R</b> adar <b>T</b> opography <b>M</b> ission
<b>SVM</b>	<b>S</b> upport <b>V</b> ector <b>M</b> achine

# **Part I**

## **Introduction**

# Living with avalanches

Ever since humans have inhabited or travelled through the Alps, they have been confronted with the dangers of snow avalanches (hereafter called avalanches). The development of settlements in mountainous regions was heavily influenced by avalanches, as for many centuries humans were not able to protect themselves effectively. In addition, heavy exploitation of the forest as well as deforestation increased the population's exposure (Mathieu et al., 2016). Almost two thirds of Switzerland are classified as mountainous regions with large portions endangered by avalanches. These regions are currently inhabited by around 2.2 million people (Egger et al., 2020) and numerous important transportation corridors lead through them. Furthermore, tourism is a very important part of the economy in the mountains, especially in winter, with up to one forth of all jobs directly or indirectly dependent on tourism (10% of the Gross Domestic Product for the cantons Grisons and Valais (GDP for 2019; Federal Statistical Office, 2024)). Consequently, a large number of people, businesses and infrastructure are exposed to and potentially affected by avalanches in Switzerland. As has been the case for centuries, avalanches still today endanger people in settlements and on transportation routes (Stethem et al., 2003; Statham et al., 2017). This is referred to a involuntary exposure, while exposure can also be voluntary when humans are deliberately moving outside of secured terrain, e.g., for ski tours (Stethem et al., 2003; Statham et al., 2017). On average, 21 people in Switzerland loose their lives in avalanches per year (past 20 years; Pielmeier et al., 2024), making it the natural hazard causing most fatalities (Badoux et al., 2016). Additionally, on average 135 avalanches per year cause damage in inhabited areas or to infrastructure (past 25 years; Harvey et al., 2013; Pielmeier et al., 2024).

The most obvious damage from avalanches is the so-called direct cost, resulting from the damage to lives or property when an object is hit (e.g., buildings, infrastructure, forests; Meyer et al., 2013). In addition, indirect costs arise from the direct damages or business interruptions (e.g., loss of customers, impedement of transport; Meyer et al., 2013). The indirect costs reach beyond the mountainous regions, affecting for example cargo businesses transporting their goods on the transportation corridors leading through the mountains. In the "avalanche winter" of 1999, the direct costs from destructive avalanches were estimated at 437 million CHF, while indirect costs were assessed at 181 million CHF (Wilhelm et al., 2000). Interestingly, the indirect costs in the tourism sector were 10 times higher than the direct costs (Nöthiger and Elsasser, 2004).



In view of the potential impact and the costs of avalanches in mountainous regions, managing the risk and mitigating damage is vitally important. This is achieved, for example, through operational services such as avalanche forecasting, hazard mapping or the installation of mitigation measures for endangered zones. These measures are continuously improved by avalanche research. Knowing avalanche history - when, where and under which conditions avalanches occurred, as well as their exact dimensions - is critical information for risk management and research. Spatially continuous knowledge on the magnitude of past avalanches over large areas, has been made possible in recent years thanks to satellite imagery. While less continuous in its coverage, webcam imagery has also proven helpful. Information extraction from these large datasets can only be handled effectively through automation, for example by combining the power of computers with advanced machine learning techniques. In light of the current exposure to avalanches, both involuntary and voluntary, and the anticipated changes in snowfall-line and precipitation patterns through climate change (e.g., Marty et al., 2023), automatically monitoring and keeping track of avalanche activity now and in the future is crucially important.

## 1.1 Research questions

This thesis explores the potential of deep learning to automatically extract information about avalanche occurrences from optical satellite imagery and from oblique webcam imagery. It addresses the following research questions:

- Can deep learning be used for the mapping of avalanches in large-scale optical satellite data, such as SPOT6/7?
- Which factors affect the agreement or disagreement between experts and how does this consequently influence the results of the automated mapping?
- What is the agreement between experts in estimating avalanche size and in mapping avalanches from oblique photographs (“traditional way”) or from aerial/space-based imagery?
- Can deep learning be used for the mapping of avalanches from webcam imagery?

The thesis is structured as follows: at the beginning, a general introduction to avalanche definitions and avalanche history (Sect. 2 and 3) is provided. An in-depth description of the applications in need of information about avalanches as well as the ways of providing such information (Sect. 4 and 5) is given next. This is followed by a detailed explanation of the background of the data and the methods employed for this thesis (Sect. 6). The research questions are explored in the following three chapters. They are based on three research articles published in renowned peer-reviewed journals within the field of snow and avalanche research (cumulative thesis in accordance with ETH provisions). The first chapter focuses on the automatic mapping of avalanches from SPOT 6/7 imagery (Chap. II), the second on the agreement of expert estimates as well as the reliability of manual labels (Chap. III) and the third on interactively segmenting avalanches from webcam imagery with deep learning (Chap. IV). The thesis concludes with a summary of the contributions of this thesis, remaining areas of research and a visionary view into the future (Chap. V).

# Classification and definition of avalanches

Avalanches can release on slopes steeper than about 30° when enough snow accumulates within a short time-span (Rudolf-Miklau et al., 2014). They can be distinguished by size or categorized into avalanche types according to the release type, the type of trigger, the form of movement or the liquid water content (EAWS, 2024b).

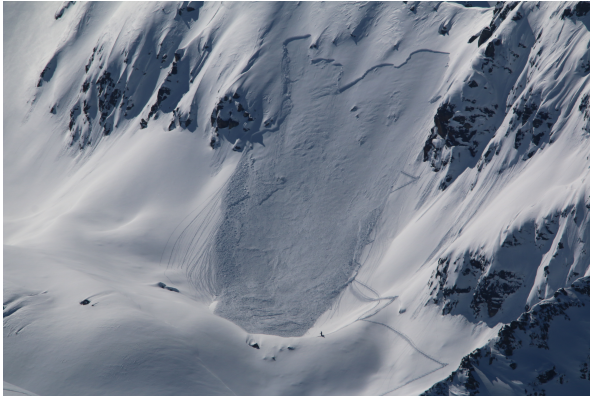
## 2.1 Avalanche types

### 2.1.1 Release type

There are different manners of starting for avalanches (McClung and Schaerer, 2006; EAWS, 2024b; SLF, 2024a): a slab avalanche may only form when there is a sufficiently widespread weak layer in the snowpack, overlaid by a well-bonded cohesive layer of snow (slab). On slopes with an inclination of 30° or steeper, the slab will slide down if a trigger (Sect. 2.1.2) causes the weak layer to collapse and the weak layer properties cause the failure to propagate over a larger area. Slab avalanches are responsible for 90% of avalanche deaths, with most fatalities being recreationists who trigger the avalanche themselves. Loose snow avalanches, in contrast, start from a point source where, due to cohesion loss, a small amount of snow starts moving downwards, entraining more snow and fanning out on the way down. Loose snow avalanches are often released naturally (Sect. 2.1.2) following a snowfall event or when significant warming takes place. Glide-snow avalanches release due to a loss of friction at the interface between the snowpack and the ground (Fees et al., 2023). The gliding movement preferably happens over smooth ground like grass slopes or smooth rock, sometimes causing glide cracks to open before an avalanche releases. Glide snow avalanches only release naturally and cannot be artificially triggered, consequently they are of minor importance to recreational activities, but significant for transportation routes.

### 2.1.2 Type of trigger

Avalanches can either occur naturally, be triggered by an additional load such as a skier, a snowmobile or an animal, or be artificially triggered using explosives (EAWS, 2024b; SLF, 2024a). In the case of spontaneous or natural avalanches, new snow or wind-transported snow deposited on the snowpack can lead to an avalanche. A rapid rise in temperatures, or a



(a) A typical slab avalanche with the sharp-edged fracture line well visible on top (photo: SOS Jakobshorn).



(b) Several loose snow avalanches starting in steep, rocky terrain and fanning out on the way down (photo: SLF/E. Hafner).



(c) Glide snow avalanche. The bare ground on which the gliding occurred is clearly visible after the release (photo: SLF/E. Hafner).

Figure 2.1: The three different avalanche types according to the manner of starting.

rain-on-snow event can weaken the snowpack and increase the likelihood of avalanches. Especially slab avalanches can be triggered by the additional load of a human, for example skiers, on top of the snowpack. The stress on the layers in the snowpack is greater when a skier jumps or falls. To secure ski areas or roads, the responsible services revert to using explosives, either from permanently installed devices, by hand or from the helicopter. The use of explosives allows for the intentional triggering of avalanches in a controlled manner.

### 2.1.3 Type of movement

The form of motion is directly influenced by the snowpack properties in the release area of an avalanche (Ancey, 2001), resulting in either a dense flow avalanche, a powder avalanche or an avalanche with a combination of those two flow regimes (EAWS, 2024b). When avalanches are released in terrain with large altitude differences, snow can be suspended in the air, forming powder avalanches reaching speeds up to 300 km/h (SLF, 2024a). In a dense flow avalanche, in contrast, the snow is primarily flowing, sliding and slipping (EAWS, 2024b) with fairly high densities between  $150 \text{ kg/m}^3$  and  $500 \text{ kg/m}^3$  (Ancey, 2001). An avalanche may include a mixture of these motion types with a dense flowing part and snow turbulently suspended (Köhler et al., 2018).

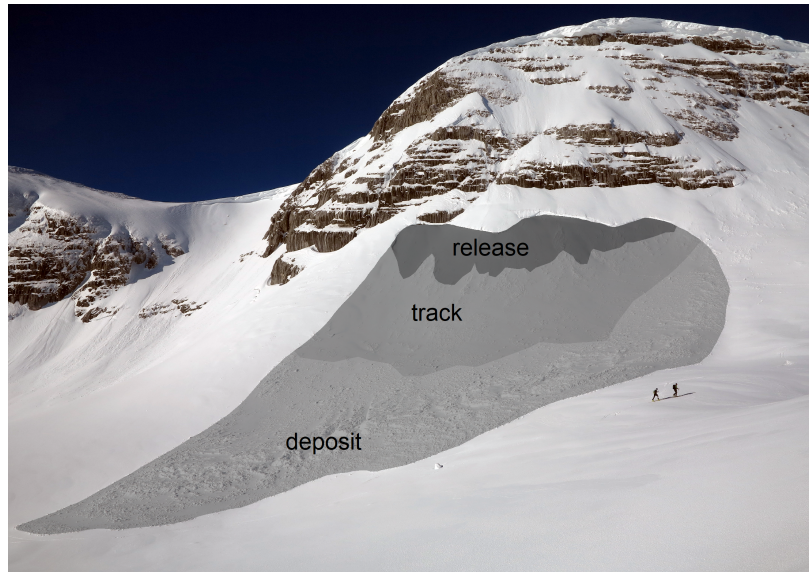


Figure 2.2: Release, track and deposit area of a slab avalanche (photo: E. Hafner).

#### **2.1.4 Liquid water content**

Depending on the presence or absence of liquid water in the snowpack, wet or dry snow avalanches will form. This classification into wet and dry snow refers to the release area of the avalanche (Sect. 2.2). Consequently, a large dry snow avalanche can arrive wet in the runout zone. The velocity of wet snow avalanches is usually lower than of dry snow avalanches. Wet snow avalanches primarily occur with the warming of the snowpack during the day, typical for springtime or rain-on-snow events. The water in the snowpack weakens the bonding at the layer boundaries, especially at layer boundaries with large differences in grain size where the water may accumulate (SLF, 2024a).






### **2.2 Avalanche parts**

Avalanches are distinguished into the release area (starting zone), the track and the deposit area (runout zone; Fig. 2.2). They start in the release area, flow down the track with the snow coming to a stop in the deposit area (EAWS, 2024b). The deposits remain visible longer than the other parts of the avalanche. Large deposits on valley floors may persist well into summer (Caiserman et al., 2022).

### **2.3 Avalanche size**

The maximum avalanche size is limited by the altitude differences of the respective mountain range and the availability of snow (Bozhinskiy and Losev, 1998). For the communication among practitioners and with the public, standardized avalanche scales have been introduced (U.S. Department of Agriculture, 1961; McClung and Schaerer, 1980). In Europe, Canada and New Zealand, among others, avalanches are classified using a scale with five size classes, with size 1 being the smallest and size 5 being the largest avalanche (EAWS, 2023). Each one of these sizes is defined by the avalanche's destructive potential, the runout/length, the volume/mass and sometimes the impact pressure (Tab. 2.1; e.g., Canadian Avalanche Association, 2016; EAWS, 2023).

Table 2.1: European definition of avalanche size (EAWS, 2023, all photos SLF, except size 4: P. Stoebener).

Size	Parameter	Definition	Example image
1: Small avalanche	Potential damage	Unlikely to bury a person, except in run out zones with unfavourable terrain features (e.g. terrain traps). Stops within steep slopes. 10-30 m 100 m <sup>3</sup>	
	Runout		
	Length Volume		
2: Medium avalanche	Potential damage	May bury, injure or kill a person.  May reach the end of the relevant steep slope. 50-200 m 1'000 m <sup>3</sup>	
	Runout		
	Length Volume		
3: Large avalanche	Potential damage	May bury and destroy cars, damage trucks, destroy small buildings and break a few trees. May cross flat terrain (well below 30°) over a distance of less than 50 m. several 100 m 10'000 m <sup>3</sup>	
	Runout		
	Length Volume		
4: Very large avalanche	Potential damage	May bury and destroy trucks and trains, may destroy fairly large buildings and small areas of forest. Crosses flat terrain (well below 30°) over a distance of more than 50 m. May reach the valley floor. 1-2 km 100'000 m <sup>3</sup>	
	Runout		
	Length Volume		
5: Extremely large avalanche	Potential damage	May devastate the landscape and has catastrophic destructive potential.  Reaches the valley floor. Largest known avalanche. >2 km >100'000 m <sup>3</sup>	
	Runout		
	Length Volume		

## Destructive avalanche periods in the past

The main driving factor for severe and disastrous avalanches is the amount and intensity of snowfall (Laternser and Pfister, 1997). Large-scale weather conditions vary between single winters, leading to periods with higher avalanche activity and more damage, sometimes referred to as “avalanche winters”. Historical records are incomplete, but existing records are often from winters with larger-than-usual avalanches that caused memorable damage. For example Peter Ruosch mentions in his avalanche chronicle of St. Anthönien, Switzerland (spanning the years 1608-1876) that in the winter of 1667/1668 ten avalanches destroyed one house, 17 large and 6 small stables, killing 55 large and 30 small livestock (Coaz, 1881). In Ftan, Switzerland Coaz (1881) reports an avalanche destroying 15 livestock and killing 36 people in February 1720. In the winter of 1950/51, a period with up to 250 cm of new snow in six days in combination with an existing weak snowpack lead to 1000 destructive avalanches causing 75 deaths and large damage in several regions in Switzerland (EISLF, 2000). Numerous large powder avalanches, formed in the winter of 1967/68 after a cold period with high wind-speeds and a lot of new snow falling on a weak snowpack. 250 avalanches were recorded to have damaged goods and infrastructure and killed 23 people (EISLF, 2000).

February of 1999 is the most recent, most widespread and longest avalanche period known in the Alps with approx. 400 to 750cm of new snow and numerous large and extremely large avalanches (Wilhelm et al., 2000). In Switzerland alone 28 people in inhabited areas or on roads were caught in avalanches. Ultimately, 17 of the victims died (Wilhelm et al., 2000). 27 buildings were partially demolished, 11 buildings were completely destroyed and around 1'700 damaged buildings were reported to the cantonal insurance companies causing an overall economic losses of more than 600 million CHF (Wilhelm et al., 2000). “Avalanche winters” and extreme avalanches often led to advancements in avalanche understanding and protection techniques as well as increased awareness and provision of financial means. After an avalanche in 1867 above Martina, Switzerland, about half a kilometer of stone walls were erected to protect the village. This is the first known utilization of stone walls in the potential release area of an avalanche (Coaz, 1881).

Periods of deforestation can historically be linked to population rise and the increased demand for wood during industrialization (Mather et al., 1999; Conedera et al., 2017). In the mid 19th century, increasing awareness of the protective effect of mountain forests against natural hazards contributed to afforestation, stricter laws, and adapted management of mountain forests (Mather and Fairbairn, 2000). In Switzerland, legislation to keep and maintain healthy forests as protection against avalanches and other natural hazards was passed and compliance enforced upon the initiative of the first Swiss forest inspector J.F. Coaz (Coaz, 1881). Afforestation led to a stark increase in forest cover: in the Swiss Alps these secondary forests, that were established after 1880, constitute approximately 43% of the total forest area nowadays (Bebi et al., 2017). These post-1880 forests are situated primarily on steep slopes, mitigating the hazard from those potential avalanche release areas ( $> 30^\circ$ ; Bebi et al., 2017). The “avalanche winters” of 1951 and 1954 brought about a major rethink in structural avalanche protection: previously used vertical earth or stone terraces were replaced by supporting structures made of steel, rope wires or wood erected perpendicular to the slope, being structures that do not fill with snow so easily and hence are more effective (Rudolf-Miklau et al., 2014). The “avalanche winter” of 1999 was thoroughly analyzed by EISLF (2000) and several aspects requiring action were identified. For example, a systematic check of the hazard maps, and their revision where necessary, was introduced (Sec. 4.3). By 2018, 98% of all necessary maps had been created in Switzerland (Bründl et al., 2019). Since 1999, the Federal Office of the Environment (FOEN) has played a significant role in financing the avalanche warning service (Sect. 4.1; Bründl et al., 2019). Additionally, the decision base for the avalanche warning service was improved by increasing the number of meteorological measurement stations from 103 in 1999 to 189 in 2023 (SLF, 2024b). As one means of improving the avalanche protection of transportation infrastructure, the number of stationary avalanche blasting systems has increased tenfold since 1999 (currently around 250; Bründl et al., 2019).

# Applications depending on proper avalanche information

There are numerous applications such as hazard mapping, mitigation measure planning and evaluation, risk analysis, avalanche warning, avalanche models and various avalanche research topics which depend on information regarding past avalanches. Generally, data on avalanche occurrences is often limited to selected locations and events. All applications would profit from a more comprehensive database of well-documented avalanche occurrences for informed decision making. The required information per application varies, such as the time gap between avalanche occurrence and the provision of information, or the specifics regarding avalanche size (Tab. 4.1).

## 4.1 Avalanche warning service

Most national avalanche warning services in Europe were established in the 20th century, forecasting the avalanche danger based on an assessment of current conditions in combination with the meteorological forecast (McClung and Schaerer, 2006; Müller et al., 2023). Current conditions are determined by meteorological measurements, observations from the field such as whumpf-sounds, shooting cracks, snow profiles, snow stability tests and the number of observed avalanches within the last 24 h (McClung and Schaerer, 2006; Techel et al., 2020). Knowing about these avalanches, their approximate location, their size, their manner of starting as well as their trigger type as soon as possible after their release is essential. In 1994, the European danger scale with five danger levels was introduced to homogenize the forecast between the different countries. There are five danger levels that might be forecasted: 5 – very high, 4 – high, 3 – considerable, 2 – moderate, 1 – low (EAWS, 2024a). They are a function of snowpack stability, the frequency distribution of that stability and avalanche size for a given unit (area and time; EAWS, 2024a). In Switzerland, the avalanche forecast is published daily at 5 pm during the winter and updated in the morning at 8 am if danger levels are at or above considerable (Fig. 4.1). For avalanche forecasting, near-real-time data availability is essential, except when retrospectively assessing forecasted avalanche risk. For this, comprehensive avalanche activity over a larger area is compared to the anticipated activity outlined in the forecast (for danger level 3 and up; EISLF, 2000; Bründl et al., 2019).



Application	Desired Input Information
Avalanche warning service	<ul style="list-style-type: none"> <li>• Information about occurrence as soon as possible (max. a couple of hours)</li> <li>• Approximate location</li> <li>• Avalanche size</li> <li>• Aspect of the release area</li> <li>• Type of trigger</li> <li>• Information on weak layer</li> <li>• Information on damage caused to e.g. people, infrastructure</li> <li>• Evaluation of forecast in hindcast: comprehensive information about avalanche occurrences (size, location), including the (approximate) date of release is essential</li> </ul>
Avalanche models	<ul style="list-style-type: none"> <li>• Date of release (time lag of reporting negligible)</li> <li>• Entire extent/ outline from release to deposit (dense part)</li> </ul>
Hazard mapping	<ul style="list-style-type: none"> <li>• Extent of the powder part</li> <li>• Extent of release area</li> </ul>
Protection/mitigation measure planning and evaluation	<ul style="list-style-type: none"> <li>• Type of trigger</li> <li>• Fracture depth</li> <li>• Mass balance/volume (distribution patterns) for the whole avalanche area</li> </ul>
Risk analysis and safety concepts	<ul style="list-style-type: none"> <li>• Information on damage caused to e.g. infrastructure, vegetation</li> </ul>

Table 4.1: Overview of applications relying on information from past avalanches, including a description of the desired input.

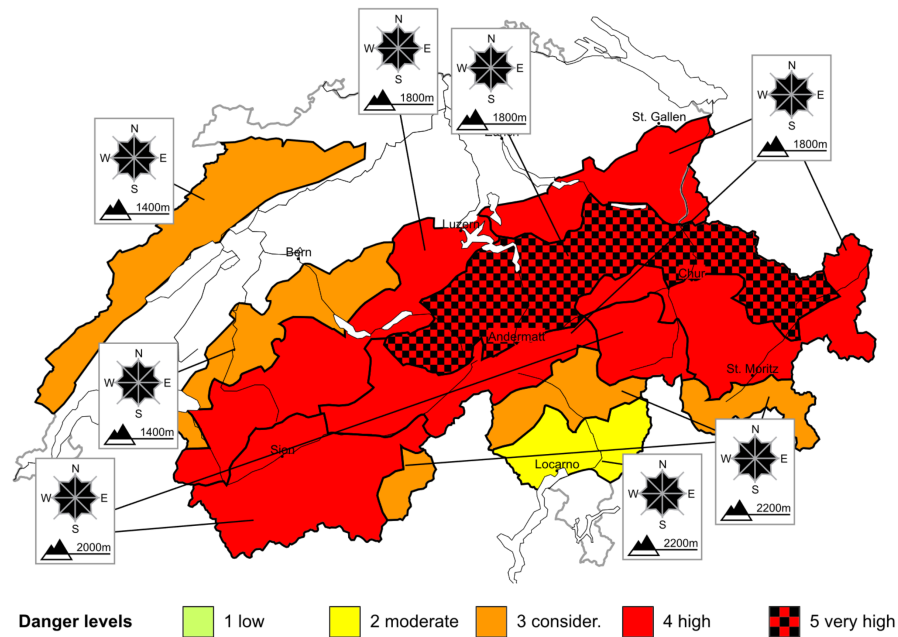


Figure 4.1: Overview of an avalanche forecast, including the danger level, the aspect and the altitude, for Switzerland (danger levels from the 14.01.2019 at 8 am, source: archive SLF).

## 4.2 Avalanche modeling

From the beginning, avalanche science attempted to understand and mathematically describe the processes underlying avalanches. The work of Voellmy (1955) on the destructive potential of avalanches is considered a milestone for avalanche modeling. Research was conducted in different places, occurring simultaneously in the former USSR (Bozhinskiy and Losev, 1998) and in the so-called West. Before the advent of computers, simple avalanche dynamics models based on steady state flow without motion tracking were used and lateral spreading was accounted for in a simplistic approach (Salm, 2004). Later, several numerical avalanche simulation software solutions like SAMOS (Sampl and Zwinger, 2004), RAMMS (Fig. 4.2, Christen et al., 2010) or AvaFrame (Tonnel et al., 2023) were developed. Avalanche hazard mapping is the primary use and motivation for avalanche dynamics models (Salm, 2004). To accurately depict reality, numerical models must undergo calibration using data from numerous field-observed avalanches across diverse snow conditions, topographies, and climates. This works best if the complete avalanche outlines, the release area, the fracture depth and the snow conditions are known and information on the mass balance/volume as well as the damage caused is available (if applicable; Salm, 2004).

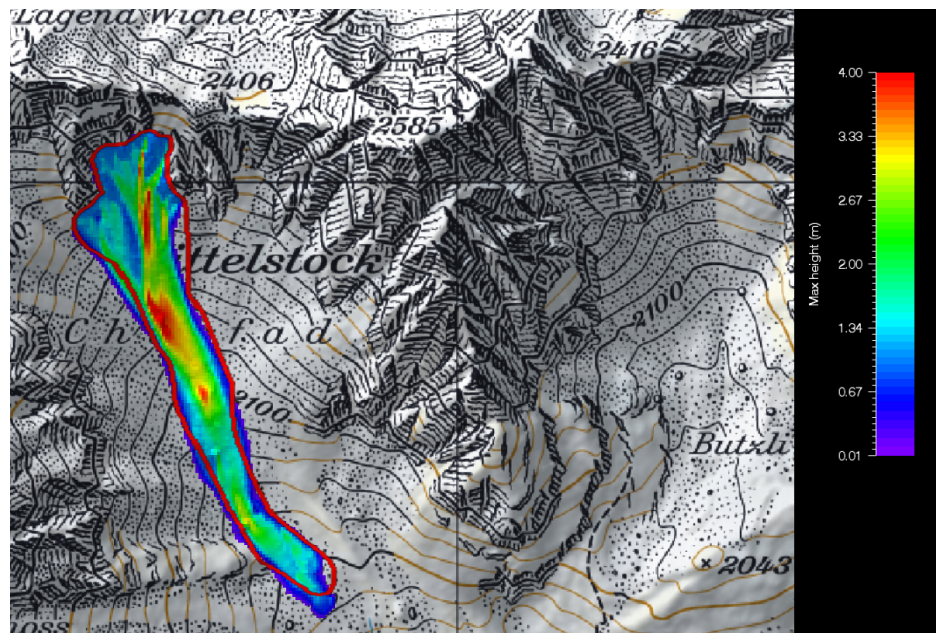


Figure 4.2: Example of an avalanche mapped from satellite imagery (red line, Hafner and Bühler, 2019), reconstructed and modeled with RAMMS::Avalanche (Christen et al., 2010, map source: Federal Office of Topography).

### 4.3 Hazard mapping

The spatial distribution of damage potential is heavily influenced by the historical growth of settlements on the one hand and spatial planning on the other hand. Preventing building activity in endangered areas through land-use planning is regarded as a cost-effective method to mitigate the risk of avalanches (Bründl and Margreth, 2021). One tool for land-use planning are hazard maps (Fig. 4.3), which indicate the endangered area, typically based on the frequency and intensity of events (Gruber and Margreth, 2001). These hazard maps delineate various levels of endangerment, which are then integrated in the communal zoning maps. These zones dictate where construction is prohibited or permitted only under specific conditions. In Switzerland (BFF and EISLF, 1984), the different zones are based on the expected avalanche intensity and frequency: red zones indicate significant danger, where there is endangerment even within buildings; in blue zones, people inside reinforced buildings are relatively safe while people are endangered outside of the buildings; in the yellow zone, there is low danger; in the yellow-white zone, some residual danger remains; while in the white zone, the avalanche danger is negligible (Fig. 4.3).

Hazard maps have a long history. The first one was established in 1954 in Gadmen, Bernese Oberland, Switzerland based on a “cadastre” of all observed avalanche runouts and expert judgment (Salm, 2004). To predict the extent and intensity of rare and possibly extreme avalanches, which might have never been observed, avalanche dynamics models were established (Salm, 2004, Sect. 4.2). Especially for areas with few or no avalanche observations, like new transportation routes or settlements built in remote areas, avalanche dynamics models have an enormous value. During the last catastrophic avalanche winter in Switzerland in 1999, 97% of all hazard maps worked as expected and the failure of the other 3% was mostly not caused by calculation errors (Gruber and Margreth, 2001), but by unforeseen scenarios such as multiple avalanches within a short period of time or unexpectedly long run-out distances of

fluidized avalanches (Wilhelm et al., 2000). Hazardous areas undergo reassessment approximately every 10 to 15 years, taking into account new insights from avalanche occurrences or changes in climate, alterations in the topography, advancements in simulation models or the implementation of new mitigation measures (Margreth, 2018).

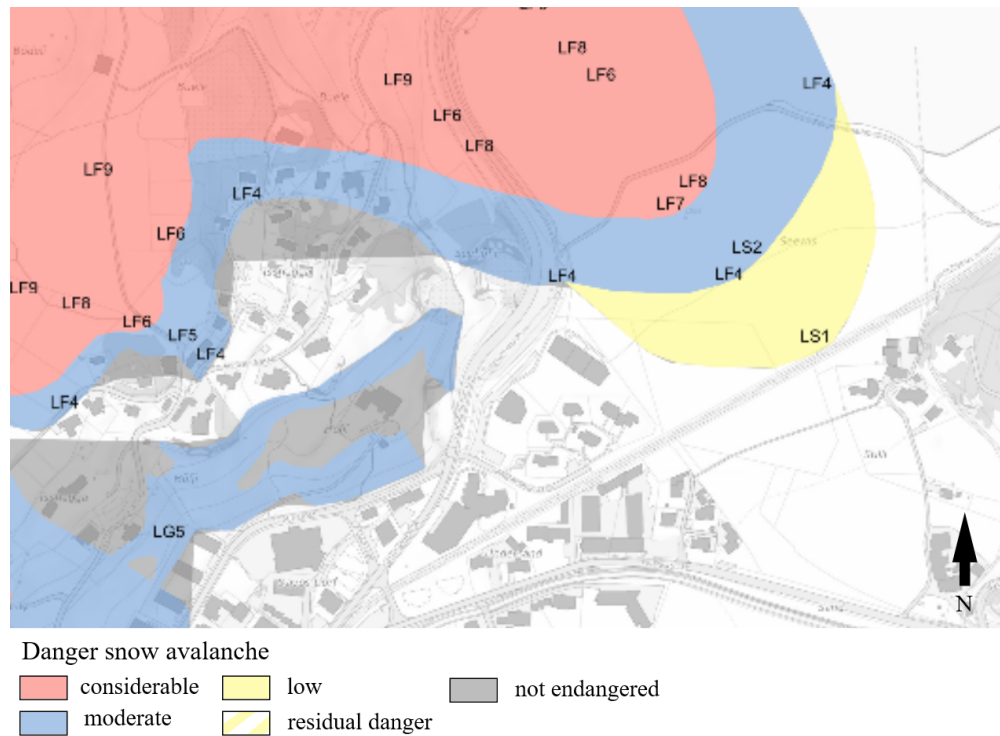


Figure 4.3: Example of a hazard map showing zones with different exposure to snow avalanches in Davos, Canton Grisons (©Amt für Wald und Naturgefahren, GR 2024, Map source: Federal Office for Topography).

#### 4.4 Planning and evaluation of protection/mitigation measures

Risk may be reduced by protection measures in the release area, on the track or in the runout zone of an avalanche: mitigation measures can be implemented to prevent the release of avalanches, influence their direction of movement or trap and retain them in a deposit area (Bründl and Margreth, 2021). In the release area, avalanche initiation may be prevented by the presence of permanent snow supporting structures (e.g., steel bridges, Fig. 4.4b) designed for a lifespan of 100 years, or temporary wooden supporting structures designed to last approximately 30 years (Bründl and Margreth, 2021). Protective forests have a similar function and reforestation up to the tree line is often applied together with temporary supporting structures until the trees are large enough. The avalanche flow direction can be influenced with deflecting structures such as dams, walls or snow sheds, while catching dams or breaking mounds reduce the avalanche velocity and retard or stop dense flow avalanches (Bründl and Margreth, 2021). In areas where the anticipated avalanche impact loads are less than 30 kPa, buildings in Switzerland are permitted to be constructed with structural enhancements, such as reinforced walls without openings in upstream direction or splitting wedges which deviate the moving snow masses on both sides of a building (Fig. 4.4a; Holub and Hübl, 2008). It is essential to carefully quantify and analyze the impact of mitigation measures in the 30-, 100-, and 300-year scenarios, as well as in an extreme scenario, surpassing the applied design scenario, in order to determine their effect on the extent of the hazard zones (Margreth and Romang, 2010). Overall, the amount

of risk reduction should ideally be greater than the amount of money required to achieve said reduction (Bründl and Margreth, 2021). Knowing the exact extent of past avalanches is important for the planning of mitigation measures, e.g. in order to identify the relevant release areas for support structures. In order to evaluate the effectiveness of existing structures, it is key to analyze the area protected by the measures as well as the surrounding area under extreme conditions, as was done for the “avalanche winter” of 1999 (EISLF, 2000).



(a) The church in Davos Frauenkirch was rebuilt with a wedge in 1602, intended to deviate future avalanches past the building on both sides. The modern building (background right) is situated in the blue zone and has a reinforced concrete wall in upstream direction without (large) openings (photo: E. Hafner).

(b) Steel supporting structures at the Grüniberg stabilizing the snowpack and preventing avalanches from starting in the release area, thereby protecting the city of Davos beneath. The structures are dimensioned according to the expected snow heights, since snowed-in structures lose their protective function (photo: I. Rittmeyer).

Figure 4.4: Examples of both historic and modern avalanche mitigation measures.

## 4.5 Risk analysis and safety concepts

The long-term avalanche risk for permanent settlements and critical infrastructure is kept low by building in compliance with hazard maps (Statham et al., 2017). However, this does not encompass short-term peaks with high avalanche risk in ski areas, backcountry areas, on transportation corridors, or other outdoor areas where people move outside of safe buildings. In these locations, it is important to assess the current risk level, defined as “the combination of the probability of an event and its negative consequences” (UNISDR, 2009), to ascertain whether it is deemed acceptable by society or not (Bründl, 2013). Short-term risk for ski resorts, temporary worksites, and transportation corridors is dealt with by safety services through the implementation of closures and controlled artificial avalanche releases (Statham et al., 2017). Local safety authorities implement warnings, closures, and evacuations for areas at risk, according to their judgement of the current situation (Bründl and Margreth, 2021). Mountain guides use professional route selection to manage the exposure of their clients, and the public avalanche forecast communicates the regional avalanche danger to an audience managing their own risk for snow-sport activities outside secured areas (Statham et al., 2017). For risk analysis, the frequency of avalanches is essential. Frequency is influenced by the climatic region, the elevation, the location of a mountain range compared to storm tracks, the location (of the object to protect) within the avalanche path and whether avalanches are artificially released, for example with explosives (Rudolf-Miklau et al., 2014). The frequency of avalanches is generally greater in the upper, steeper sections of the avalanche path (Rudolf-Miklau et al., 2014). Risk analysis is based on scenarios characterized by the frequency, the spatial extent and the physical

impact (intensity) of avalanches, the most important basis for scenarios being data from past events (Bründl, 2013). Nearest neighbour methods might be used to search for past situations with similar conditions, assuming a similar avalanche activity (e.g., Kristensen and Larsson, 1994).

# Documentation of avalanches

## 5.1 Historical events and data from the past

Gathering and analyzing information on documented historical avalanches is the initial source of knowledge for understanding their spatial and temporal extent (Bonnefoy et al., 2012). Long time series enable the understanding of the dynamics (return period) and the definition of reference scenarios (Giacona et al., 2017). Generally, the information on past events is denser near inhabited areas where impacts on life and personal goods can be observed by a large(r) number of people (Bonnefoy et al., 2012). However, the information is usually scattered across various historical sources, varies in detail and is only available for selected avalanches (Butler, 1986). For example, information on historic avalanches (between 1910–1982) was compiled in the Glacier National Park, US from ranger logbooks, annual park reports, photographs, local weekly newspapers and state highway reports. Giacona et al. (2017) reconstructed 730 avalanches with historical and geographical methods between the winters of 1783/1784 and 2013/2014 on 128 paths in the Vosges mountains in France. If trees were affected by the avalanches, activity can be reconstructed with tree-ring analysis, even if no observers were present (e.g., de Bouchard d'Aubeterre et al., 2019). Systematic large-scale recordings of avalanches (with records still available) through observers are known from the beginning of the second half of the 19th century for certain sections of the Georgian Military Road (Bozhinskiy and Losev, 1998) or for the whole canton of Grisons in Switzerland for the year of 1872 through the forest service (Coaz, 1881).

## 5.2 Avalanche documentation and mapping today

Numerous safety related applications (Sect. 4) as well as research are dependent on knowing about past avalanches to continuously ensure safety and advance avalanche understanding. Currently, there are several possibilities to record avalanches: ground based (“traditionally”, through observers), with seismics, with infrasound, with ground-based radar, from webcam imagery or space/aerial-based like satellite, airplane or drone data. The requirements per application (Tab. 4.1) differ and depending on the way the avalanches are documented, additional information can be recorded and provided (Tab. 5.1). Especially for the sensors generating lots of data, the mere capability of capturing avalanches is not enough as manual interpretation of the data is too time consuming. Consequently, not only the sensors, but also the processing determines if the data is of value for the applications dependent on it (Sect. 4).

### **5.2.1 Observers**

Residents in mountainous areas or individuals working in these regions regularly report avalanches, supplemented by occasional reports from recreational users (Jamieson et al., 2009; Kosberg et al., 2013; Techel et al., 2022). In Switzerland, there are about 80 observers reporting avalanches in their “stationary” region on a daily basis (Techel et al., 2022), including information on the location, the estimated time of the release, the avalanche size (size classes 1 to 5 according to EAWS (2023)), the moisture content (dry or wet), the trigger type (i.e., natural release, human-triggered) and the damage caused (if applicable). Especially for larger avalanches and those close to infrastructure, the complete outlines are often mapped as well, either directly in the field or later from photographs.

### **5.2.2 Seismic sensors**

The ground motion caused by avalanches is strong enough to be recorded by seismometers or geophones. This was first investigated by Harrison (1976) about 50 years ago. The detection range of a seismic monitoring system is limited to approximately 2–4 km around the sensor (Heck et al., 2018b; Steinkogler et al., 2018). The avalanche size can be estimated from the seismic signals (Pérez- Guillén et al., 2016) as well as the rough direction of the path the avalanche followed, given an array of sensors is used (Heck et al., 2018b). The sensors are independent of weather conditions and may be used to monitor avalanche activity continuously in selected regions (Heck et al., 2018a; Steinkogler et al., 2018).

### **5.2.3 Infrasound**

The flowing and turbulent snow masses of an avalanche create pressure waves in the air that can be measured by infrasound sensors (Bedard Jr et al., 1988; Kogelnig et al., 2011; Mayer et al., 2020). Infrasound is also strongly affected by noise produced by natural (wind, earthquakes) and artificial sources (planes, helicopters, industry), which may be countered by using an array of sensors keeping the signal-to-noise ratio low (Schimmel et al., 2017). Using an array allows for the avalanche detection to be assigned to a certain avalanche path up to a distance of approx. 3 km, given the absence of major topographic barriers (Schimmel et al., 2017; Mayer et al., 2020). The amplitude of the recorded infrasound signal is controlled by the avalanche type and the flow evolution allowing an approximate reconstruction of the avalanche size (Marchetti et al., 2020).

### **5.2.4 Doppler radar**

Doppler radars can be used to monitor selected avalanche paths/ slopes by emitting electromagnetic radiation in the range of a few to several tens of GHz to a target area (Meier et al., 2016). The beam is consequently reflected, returned to the antenna and recorded by the receiver. If an object within the target area is moving towards the radar, the reflected signal will experience a frequency shift enabling avalanche detection (Schimmel et al., 2017; Meier et al., 2016). With a Doppler radar, velocities between 1 and 300 km/h can be measured in real time (Koschuch, 2018). Within their target area, radars are very reliable under all weather conditions, detecting also small avalanches without serious false alarms, if well parameterized (Schimmel et al., 2017). The length of the frequency shift gives an indication of the size of the detected avalanche.

### **5.2.5 Webcam data**

With time-lapse photography, avalanches in the camera’s field of view can be monitored. This has mostly been done for glide snow cracks and/or glide snow avalanches (Hendrikx et al., 2012; Feick et al., 2012; Helbig et al., 2015; Fees et al., 2023), wet snow avalanche activity (Helbig et al., 2015; Abermann et al., 2019) or more recently all avalanche types and glide snow



cracks (Fox et al., 2023). The monitoring of avalanche activity with automatic webcams provides accurate near-real time data, though bad weather conditions hinder continuous monitoring (Eckerstorfer et al., 2016). Thanks to the high temporal resolution of many webcams, the likelihood of capturing an avalanche is higher than with the “single-shot” of satellite data acquisition. However, compared to airborne or satellite-borne optical remote sensing, the spatial coverage is limited (Eckerstorfer et al., 2016).

### **5.2.6 Satellite data**

Satellite data covers terrain otherwise inaccessible (Eckerstorfer and Malnes, 2015; Eckerstorfer et al., 2016) and may be used to document avalanches over large regions (Bühler et al., 2019; Eckerstorfer et al., 2019). For avalanches, both passive optical sensors, capturing the visible light, as well as active sensors using microwaves are in use:

#### **SAR**

Wiesmann et al. (2001) were the first to describe the potential of mapping avalanches from active synthetic aperture radar (SAR). Avalanches may be distinguished from surrounding undisturbed snow by differences in backscatter (Wiesmann et al., 2001; Eckerstorfer and Malnes, 2015) or with change detection (Vickers et al., 2016; Leinss et al., 2020; Eckerstorfer et al., 2019; Hafner et al., 2021; Keskinen et al., 2022; Eckerstorfer et al., 2022) under all-weather and all light conditions (Eckerstorfer et al., 2016). The backscatter changes are strongest in the deposit area of avalanches, limiting detection of the other avalanche parts (Vickers et al., 2016; Hafner et al., 2021; Keskinen et al., 2022). While most current work relies on free of charge Sentinel-1 data (Vickers et al., 2016; Wesselink et al., 2017; Abermann et al., 2019; Eckerstorfer et al., 2019; Leinss et al., 2020; Hafner et al., 2021; Eckerstorfer et al., 2022; Kapper et al., 2023), various sensors have been proposed and tested in the last approx. 25 years: ERS1/2 (Wiesmann et al., 2001), Radarsat-2 Ultrafine (Eckerstorfer and Malnes, 2015; Wesselink et al., 2017) and TerraSAR-X (Bühler et al., 2019; Leinss et al., 2020). With Sentinel-1 imagery, an overall Probability of Detection (POD; see Equation 11.2) between 11.3% (for exclusively dry snow avalanches; Eckerstorfer et al., 2022) and 27% (for a mixture of dry and wet snow avalanches; Hafner et al., 2021) was achieved. The performance is better for larger avalanches, for example for size 3 avalanches (EAWS, 2023) the POD was found to be 42% (Hafner et al., 2021) and 65% (Keskinen et al., 2022) respectively.

#### **Optical**

Images from several optical satellites with varying resolution have been tested for avalanche mapping: SPOT-2 and SPOT-4 (Satellite Pour l’Observation de la Terre; 20 and 10 m spatial resolution; Frauenfelder et al., 2012), Quickbird (0.65 m; Frauenfelder et al., 2012; Lato et al., 2012), WorldView-1 (0.5 m; Feick et al., 2012; Frauenfelder et al., 2015), Landsat-8 (15 m; Eckerstorfer and Malnes, 2015), World-View 4, Pleiades, SPOT 6/7 (0.3 m, 0.5 m, 1.5 m; Bühler et al., 2019) and Sentinel-2 (10 m; Abermann et al., 2019; Hafner et al., 2021). For the large scale mapping of avalanches in Switzerland, Bühler et al. (2019) found SPOT 6/7 to be the best compromise between large area coverage as well as good avalanche detectability and mapped over 18’000 avalanches in 12’500 km<sup>2</sup> of imagery. Hafner et al. (2021) showed that the mapping from SPOT 6/7 captures 74% of all avalanches ( $\geq$  size 2) and 91% of size 3 avalanches (EAWS, 2023, POD scores given). The maximum resolution for avalanche identification depends on the avalanche type: Abermann et al. (2019) was able to identify very wet avalanches (slush flows) in Greenland from Sentinel-2 with a relatively coarse resolution of 10×10m, while Hafner et al. (2021) found Sentinel-2 unsuitable for avalanche identification for two periods with dry and wet snow avalanches in Switzerland. As opposed to SAR, the whole avalanche outline can be identified from the imagery, at least in well illuminated areas. The largest limitation of optical

satellite data is the dependency on cloud and fog-free weather conditions during data acquisition, the cast shadow as well as the absence of solar illumination in higher latitudes during winter (Frauenfelder et al., 2012; Eckerstorfer et al., 2016; Bühler et al., 2019).

### **5.2.7 Airplane and Drone data**

Airborne optical remote sensing data have been proposed for avalanche mapping relying on different sensors with a spatial resolution of approx. 25 cm (Bühler et al., 2009; Frauenfelder et al., 2012; Lato et al., 2012; Korzeniowska et al., 2017; Bührle et al., 2023). Drone imagery has been tested for the documentation of selected avalanches, e.g. in roadside avalanche monitoring (McCormack and Vaa, 2019). High-resolution orthophotos allow for avalanche identification from the release area to the deposit area. In addition, a Digital Surface Model (DSM) may be calculated with photogrammetry as shown by Bührle et al. (2023), who have shown this with a Root Mean Square Error (RMSE) of 0.1 to 0.2 m from airborne data while Bühler et al. (2016, 2017) demonstrated a RMSE of 0.01 m for drone imagery. Consequently, the fracture depth, volume, mass balance and snow distribution for and around each avalanche can be derived with high accuracy (Bühler et al., 2017; Bührle et al., 2023), helping to understand flow and deposition patterns. Even though airplane and drone data are dependent on good weather conditions, as any optical data, they may acquire data when flying under high clouds. However, especially drones might be additionally restricted in operation by high wind speeds.



# Data and processing methods

## 6.1 Sensors and datasets

This thesis uses several remotely-sensed datasets, from Open Government (OGD) Swiss datasets from the Federal Office for Topography (swisstopo), commercial Satellite Pour l'Observation de la Terre (SPOT) satellites to open access data from the Copernicus Sentinel missions. The following sections give a background on the characteristics of each sensor and dataset of importance for this thesis.

### 6.1.1 Sentinel-1

The Sentinel-1 mission has a constellation of two sun-synchronous polar-orbiting satellites both equipped with a C-band Synthetic Aperture Radar (SAR) instrument, an active sensor (center frequency 5.405 GHz; ESA, 2024b). Sentinel-1A was launched in 2014 and is still in operation, while Sentinel-1B was launched in 2016 and failed in December 2021. With two satellites in operation, the orbit repeat interval at mid-latitudes like Switzerland was six days, with one satellite it doubled to 12 days. Sentinel-1C is planned to launch in 2024, bringing the orbit repeat interval back to 6 days. The satellites possess four different imaging modes, with the Interferometric Wide (IW) swath mode being the main acquisition mode over land ( $5 \times 20$  m spatial resolution). The SAR antenna measures returning amplitude and phase information from the emitted microwave signal. The system may acquire data under all weather conditions and independent of daylight, a significant advantage over all optical sensors. The data is publicly available and free of charge.

### 6.1.2 Sentinel-2

The Sentinel-2 mission collects high-resolution multi-spectral images utilizing two identical satellites in sun-synchronous orbits: Sentinel-2A and Sentinel-2B (ESA, 2024b). Together they result in an orbit repeat interval of 6 days at mid-latitudes. The Multi Spectral Instrument (MSI) acquires data in 13 spectral bands ranging from the visible bands to the near and short wave Infrared (NIR and SWIR). The Red, Green, Blue (RGB) and NIR are acquired with a spatial resolution of 10 m, while the other Infrared bands have 20 m resolution and the bands focused on cloud screening and atmospheric correction have 60 m resolution. Like Sentinel-1, the data is publicly available and free of charge, but usable data is only acquired under cloud-free weather conditions.

### 6.1.3 SPOT 6/7

SPOT 6 and SPOT 7 are the last satellites from the SPOT series owned by Airbus Space and Defense, a series originally conceived by the French National Centre for Space Studies (CNES). The identical satellites were launched in 2012 and 2014 respectively, with SPOT 6 being still in operation while SPOT 7 went out of operation in March 2023 (ESA, 2024a). They acquire optical imagery from a sun-synchronous orbit at an altitude of 694 km. With two New Astrosat Optical Modular Instruments (NAOMI) onboard, SPOT 6/7 can achieve a swath of 120 km<sup>2</sup> and consequently capture large areas with one overpass. At nadir data is acquired with 1.5 to 2.5 m Ground Sample Distance (GSD) in the panchromatic channel and 6 to 10 m GSD in the multispectral channels (RGB, NIR). After pan-sharpening, the multispectral channels are usually provided with 1.5 m spatial resolution. As a constellation, the satellites had a orbit repeat interval capability of one day, with only one satellite the orbit repeat interval capability is between one and three days. The satellites are commercial, the cost is approx. 4 CHF per km<sup>2</sup>, but it might be higher if the data is prioritized and ordered in Rapid Mapping mode.

### 6.1.4 Ultracam airplane imagery

The Voxel UltraCam Eagle M3 is an airplane-mounted camera for photogrammetric measurements with 450 megapixel (Bührle et al., 2023). It acquires data in the four spectral bands RGB and NIR with a radiometric resolution of 14 bits and a GSD of 12 cm at a flight height of approx. 4000 m a.s.l.. The camera positions are tracked using the differential Global Navigation Satellite System (dGNSS) with a nominal accuracy of 0.2 m. The camera orientation is recorded with an inertial measurement unit (IMU) possessing a nominal accuracy of 0.01° (omega, phi, kappa). To ensure a positional accuracy of approx. 0.1 m, Ground Control Points (GCPs) are used. The orthophotos are generated with Agisoft Metashape (structure-from-motion) processed to a spatial resolution of 0.25 m and the digital surface model (DSM) with 0.5 m spatial resolution.

### 6.1.5 Wingtra drone imagery

The Wingtra One is a fixed-wing drone able to cover about 5 km<sup>2</sup> of high alpine terrain on a daily basis (Wingtra, 2024). The drone is equipped with Post-Processing Kinematics (PPK) and acquires data in RGB with a Sony RX1R II 42 megapixel camera achieving a GSD of 4 cm at a flight altitude of roughly 150 m a.s.l.. The orthophotos are processed with Agisoft Metashape to 4 cm and the DSM to 10 cm spatial resolution.

### 6.1.6 Swiss national DEM

swissALTI<sup>3D</sup> is the precise Swiss digital elevation model (DEM), representing the surface of Switzerland without vegetation and buildings with a six year repeat cycle (swisstopo, 2018). The data is based on a combination of sources: Light Detection and Ranging (LiDAR) and photogrammetry. The datasets are freely available at a spatial resolution of 0.5 m, 2 m, 5 m and 10 m. The quality varies due to the combination of different methods but the accuracy is between 30 and 50 cm below 2000 m a.s.l. and between 1 and 3 m above.

### 6.1.7 SWISSIMAGE

SWISSIMAGE is the Swiss summer orthophoto mosaic, based on color aerial photographs, with a spatial resolution of 10 cm, except for the Alps where the resolution is coarser with 25 cm. The positional accuracy lies between 0.1 and 0.25 cm (swisstopo, 2020a). The images are updated every three years with an ADS100 camera from Leica Geosystems mounted on a plane. The imagery acquired in the NIR and the RGB channels are orthorectified with the swissALTI<sup>3D</sup>. To generate a homogeneous mosaic, the images are radiometrically corrected and artefacts, like

bridges, are manually corrected.

### 6.1.8 Swiss national topographic map

Swiss national topographic maps encompass all digital national maps in raster format (1:10'000 to 1:1 million; swisstopo, 2020b). The data are derived from the digital map basics and are available as georeferenced rasters. The content is identical to the printed national maps and is often used as a background information or as basemaps.

## 6.2 Semantic segmentation with neural networks

Computer vision includes a set of techniques enabling machines to interpret and understand visual information from the world, similar to the way humans perceive and process visual information (Gu et al., 2018). Semantic segmentation is a fundamental computer vision task that aims at classifying each pixel in an image into a specific class or category. In the past years semantic segmentation shifted focus from traditional machine learning algorithms to convolutional neural networks (CNNs; Sultana et al., 2020).

Most deep network architectures for segmentation are based on the encoder-decoder paradigm (Fig. 6.1): The encoder network takes the input and through a layer of learnable filters extracts features from the input image at different levels. The decoder network takes the condensed information at different layers and through a series of upsampling/deconvolutional layers increases their size to the original size of the input image.

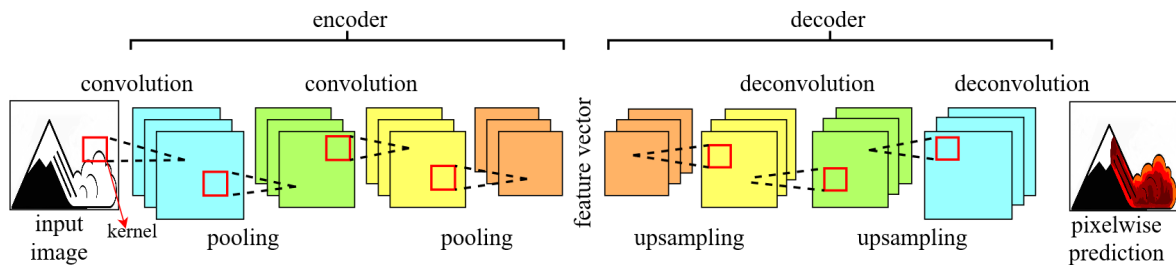


Figure 6.1: Overview of the basic architecture of a CNN.

At the core of the encoder are multiple convolutions, mathematical operations intended to extract the local information from the input images. For each convolutional layer several convolutional kernels, small matrices of learnable weights, typically square, are used to compute different feature maps (Goodfellow et al., 2016; Gu et al., 2018). Like sliding windows, the kernels look at one neighborhood before moving to the next, until they have covered the whole image and eventually the whole dataset (Goodfellow et al., 2016). The weights of the kernels are learned while training. The first layers of convolutions may identify and learn edges, textures and shapes, while later the information they extract becomes more complex and abstract (LeCun et al., 1998). Computational efficiency is achieved by the sliding of the kernels, each with a small set of weights (Goodfellow et al., 2016). After each convolutional operation, an activation function is applied element-wise to the output followed by a pooling operation or another convolution. Non-linear activation functions, such as Rectified Linear Unit (ReLU), enable CNNs to learn complex non-linear relationships and representations in the data. Pooling layers downsample the feature maps obtained from the convolutional layers, reducing their spatial dimensions and increasing the receptive field while retaining important information. Typical pooling operations would be max pooling (Boureau et al., 2010), keeping only the maximum value from a fixed kernel or average pooling (Wang et al., 2012), retaining the average value only. The decoder combines upsampling operations, skip connections, convolutional layers, activation functions,

and regularization techniques to make detailed predictions based on the low-resolution feature maps. Upsampling is mostly done with unpooling (e.g., nearest neighbor) or deconvolutions (transpose convolution, upconvolution), increasing the spatial resolution of feature maps. The output layer is the last layer of a CNN, providing a probabilistic classification based on extracted features (Gu et al., 2018). This architecture of CNNs helps to capture and learn both local and global patterns in the input data (Goodfellow et al., 2016). For a short explanation of more CNN-related terms see Appendix A.

### **6.3 Monophotogrammetry**

Monophotogrammetry, also called mono-plotting, is a technique to derive 3D information from a single photograph. The development of models to achieve this was largely driven by the vast amount of available historical photographs and the desire to study landscape change (e.g., Kull, 2005) or to analyze natural hazards (e.g., Conedera et al., 2018). Monophotogrammetry was first proposed in the 1970ies (Conedera et al., 2018) and relates a single image to the corresponding DEM. By using Control Points (CPs), selected pixels in the image (unambiguous landmarks) are given known real world coordinates that are then used to calculate the extrinsic and intrinsic camera parameters of the image (Bozzini et al., 2012). A minimum of four CPs is required to determine the unknown camera parameters. If the CPs are well distributed in the photo and in the terrain, the algorithms should automatically move towards the best solution (Samtaney, 1999). The accuracy of the results depends on the quality of the used photograph as well as the angle of incidence in which it was taken, the CPs (number, precision, distribution), the DEM (resolution and accuracy) and the accuracy of the camera calibration (Conedera et al., 2018). With accurate data and well distributed CPs, an accuracy of less than a meter may be achieved (Conedera et al., 2018). There are several monophotogrammetry tools available for solving the task, like Bozzini et al. (2012, 2013), Produit et al. (2016) or Golparvar and Wang (2021).

# Preceding work on optical satellite avalanche mapping

## 7.1 Large-scale mapping in Switzerland

The research is built upon work conducted before this thesis, namely the large scale mapping of avalanches from optical SPOT 6/7 imagery: In the beginning of January 2018, for the first time since 1999, very high avalanche danger (level 5; EAWS, 2024a) was forecasted and confirmed. At SLF, the question arose if the avalanche activity might be captured with satellite data. Data from various sensors: optical WorldView-4, Pleiades, SPOT 6/7 and TerraSar-X radar data in SpotLight and StripMap mode was acquired and analyzed to test this hypothesis (Bühler et al., 2019).

After the analysis Bühler et al. (2019) concluded that SPOT 6/7 is the best compromise between large area coverage and a spatial resolution suitable to identify avalanches. Towards the end of January 2018, another major snowfall led to large-scale very high avalanche danger (level 5; EAWS, 2024a) being forecasted. Together with the Swiss Federal Office for the Environment (FOEN), the swisstopo ordered SPOT 6/7 data over all warning regions where level 5 was predicted (Bühler et al., 2019). The same procedure was repeated approximately a year later in January 2019, when again very high avalanche danger (level 5; EAWS, 2024a) was forecasted (Zweifel et al., 2019). From both avalanche periods cloud-free imagery covering approximately 22'000 km<sup>2</sup> was acquired and orthorectified by swisstopo. The captured avalanches were consequently manually identified and digitized over the whole area (Bühler et al., 2019; Bründl et al., 2019). For each avalanche outline (2018: 18'737; 2019: 6'041) metadata describing important attributes, like avalanche type or avalanche size as well as how well the avalanche was visible, were saved (Bühler et al., 2019).

## 7.2 Investigation into performance and completeness

The performance and reliability of mapping avalanches from the chosen sensor were consequently investigated by Hafner et al. (2021). The validation was performed relying on photographs from the ground and from helicopters as a reference for an area of approximately 180 km<sup>2</sup> around Davos, Switzerland. The performance was quantified using probability of detection (POD) and the positive predictive value (PPV; see Equation 11.2). Hafner et al. (2021)



showed that the POD for the mapping from SPOT 6/7 imagery is 0.74 and the PPV 0.88 (for avalanches  $\geq$  size 2). Larger avalanches are more likely to be mapped than small avalanches (Fig. 7.1) and avalanches located fully in the shaded slopes are missed more easily than those on illuminated slopes (Hafner et al., 2021). These analyses confirmed the great potential of SPOT 6/7 for the comprehensive mapping of avalanches. However, as noted previously by Eckerstorfer and Malnes (2015), large image datasets can only reasonably be handled using automatic avalanche detection algorithms.

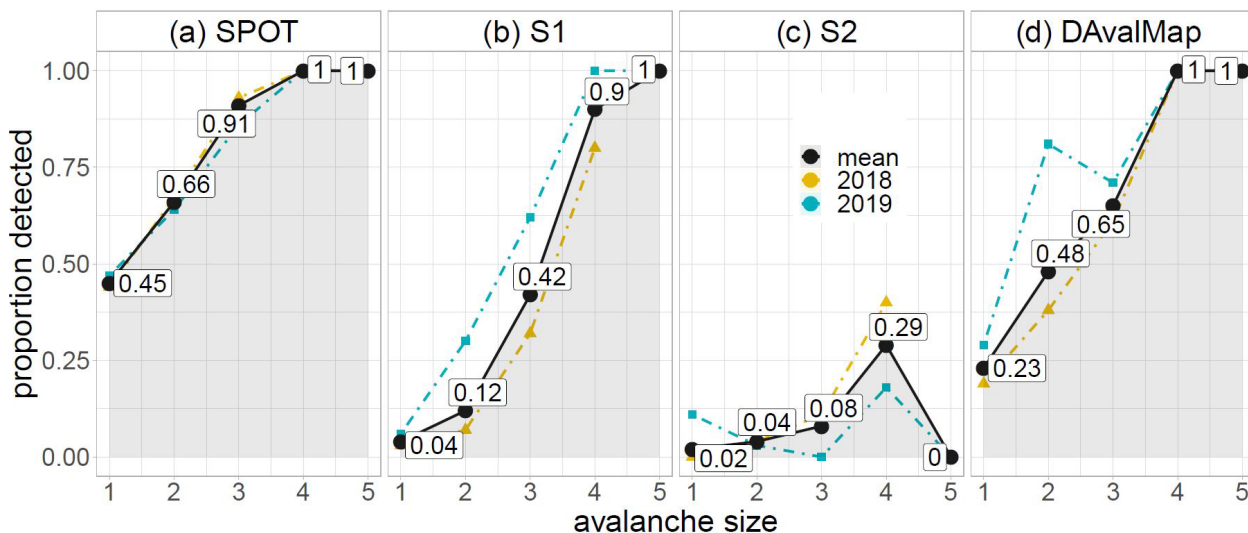


Figure 7.1: Probability of detection (POD) by size for the four avalanche mapping methods tested by Hafner et al. (2021, SPOT: SPOT 6/7, S1: Sentinel-1, S2: Sentinel-2, DAvalMap: Davos avalanche mapping project): The black dots and line represent the mean proportion of avalanches per size identified by the different mapping methods; additionally, the proportions are shaded grey. In addition, the values for 2018 (yellow triangles) and 2019 (turquoise squares) are shown for each mapping method (Fig. from Hafner et al. (2021)).

The aim of this thesis is to enlarge the databases on avalanche occurrences for the safety-related applications by using deep-learning to automatically identify and map the avalanches in the SPOT 6/7 imagery. Furthermore, this thesis pursues the goal of better understanding how reliable both the estimation of avalanche size as well as the mapping of avalanches in the “traditional way” in the field/from photographs or from remotely sensed imagery are. In this context, reliability is defined as to how consistent repeated measurements or judgments of the same event with the same process are (Cronbach, 1947). Lastly, with Interactive Avalanche Segmentation (IAS), the thesis eases the mapping of avalanches from webcam imagery, tapping into another data source for documenting avalanche occurrences.

# Part II

## Automated avalanche mapping from SPOT 6/7 satellite imagery with deep learning: results, evaluation, potential and limitations

Elisabeth D. Hafner, Patrick Barton, Rodrigo Caye Daudt, Jan Dirk Wegner, Konrad Schindler, Yves Bühler

**The Cryosphere, published: 02.09.2022**

**Author version:** for the published typeset version please refer to Hafner et al. (2022)

**Author contributions:** EDH coordinated the study, performed all initial manual mappings, expanded the neural network code originally implemented by PB and did the statistical analysis. RCD, JDW and KS advised on the machine learning aspects of the project and critically reviewed the associated results. RCD wrote the script for the Wallis filtering. The reproducibility investigation was initiated by KS and coordinated by EDH, and both YB and EDH were part of its mapping team. EDH wrote the manuscript with help from all other authors. EDH and YB originally initiated the automation of avalanche mapping from SPOT.

## **Abstract**

Spatially dense and continuous information on avalanche occurrences is crucial for numerous safety-related applications such as avalanche warning, hazard zoning, hazard mitigation measures, forestry, risk management and numerical simulations. This information is today still collected in a non-systematic way by observers in the field. Current research has explored the application of remote sensing technology to fill this information gap by providing spatially continuous information on avalanche occurrences over large regions. Previous investigations have confirmed the high potential of avalanche mapping from remotely sensed imagery to complement existing databases. Currently, the bottleneck for fast data provision from optical data is the time-consuming manual mapping. In our study we deploy a slightly adapted DeepLabV3+, a state-of-the-art deep learning model, to automatically identify and map avalanches in SPOT 6/7 imagery from 24 January 2018 and 16 January 2019. We relied on 24 778 manually annotated avalanche polygons split into geographically disjointed regions for training, validating and testing. Additionally, we investigate generalization ability by testing our best model configuration on SPOT 6/7 data from 6 January 2018 and comparing it to avalanches we manually annotated for that purpose. To assess the quality of the model results, we investigate the probability of detection (POD), the positive predictive value (PPV) and the F1 score. Additionally, we assessed the reproducibility of manually annotated avalanches in a small subset of our data. We achieved an average POD of 0.610, PPV of 0.668 and an F1 score of 0.625 in our test areas and found an F1 score in the same range for avalanche outlines annotated by different experts. Our model and approach are an important step towards a fast and comprehensive documentation of avalanche periods from optical satellite imagery in the future, complementing existing avalanche databases. This will have a large impact on safety-related applications, making mountain regions safer.

# Introduction

Information about avalanche occurrences, their location and dimensions is pivotal for many applications such as avalanche warning, hazard zoning, hazard mitigation infrastructure, forestry, risk management and numerical simulations (e.g., Meister, 1994; Rudolf-Miklau et al., 2014; Bebi et al., 2009; Bründl and Margreth, 2015; Christen et al., 2010; Bühler et al., 2022). Currently this information is reported and collected unsystematically by observers and (local) avalanche warning services. In recent years different groups have proposed to use remote sensing to fill that gap and provide spatially continuous, complete maps of avalanche occurrences over some region of interest (Bühler et al., 2009; Lato et al., 2012; Eckerstorfer et al., 2016; Korzeniowska et al., 2017). It has been shown that avalanches can be identified with sufficient reliability from optical data (e.g., Bühler et al., 2019) or synthetic aperture radar (SAR; e.g., Eckerstorfer et al., 2016; Abermann et al., 2019), with varying degrees of completeness depending on the sensor and the size of the avalanches (Hafner et al., 2021).

Both optical and SAR data have inherent advantages and disadvantages which we would like to elaborate on in the following section: for the acquisition of suitable data, SAR is independent of cloud cover, whereas for optical data a clear sky is a crucial prerequisite. Consequently, optical data may only capture avalanche occurrences after a period with activity is over (except for avalanches releasing solely due to the warming during the day), whereas with SAR information may also be retrieved during an avalanche period. Due to that independence from low-visibility weather conditions, and in the case of Sentinel-1 a 12-day repeat cycle at midlatitudes, the temporal resolution is in the best case daily in northern Norway or about every 6 days in central Europe (numbers for two Sentinel-1 satellites acquiring data, currently the temporal resolution is about half as Sentinel-1B has not been acquiring since 23 December 2021). The optical satellite data currently known to be suitable for avalanche mapping need to be ordered specifically and are therefore only available at isolated dates in time. Compared to SAR, optical data are however easier to process and interpret. In our previous work (Hafner et al., 2021) we compared the performance and completeness of SAR Sentinel-1 as well as optical SPOT 6/7 and Sentinel-2 for avalanche mapping. In a detailed analysis of the manual mappings we found the following: the ground sampling distance of 10m makes Sentinel-2 unsuitable for the mapping of avalanches. The mapping from SPOT 6/7 is overall more complete compared to Sentinel-1, which is mostly caused by the inability to confidently map avalanches of size 3 and smaller in Sentinel-1 imagery, a characteristic related to the underlying spatial resolution of approximately 10-15 m for

Sentinel-1 and 1.5 m for SPOT 6/7. Depending on the application, practitioners not only want to know when and where an avalanche occurred, but also the outlines. When analyzing which part of an avalanche can typically be identified using Sentinel-1 we found (in accordance with, among others, Eckerstorfer et al., 2022) that it is mostly the deposit, but may include patches from track and release area. When only using Sentinel-1 data it is therefore not possible to derive the number of avalanche occurrences (possibly several unconnected patches for one avalanche) or the size of the avalanche occurrences (size of patches detected does not usually correspond to avalanche size). Consequently, unless unambiguous with respect to the terrain, the origin and release area of avalanche deposits detected using SAR images remain unknown. In contrast, except for shaded areas, in SPOT 6/7 avalanches can be identified from release zone to deposit in almost all cases. Additionally, research suggests SAR to be a lot less reliable for detecting dry snow avalanches compared to wet snow avalanches (among others Hafner et al., 2021; Eckerstorfer et al., 2022). The above statements made about SPOT 6/7 are transferable to optical data with similar or better spatial and spectral resolution.

To bypass the time-consuming manual mapping, several groups have explored (semi-) automatic mapping approaches. Bühler et al. (2009) used a processing chain that relies on directional, textural and spectral information to automatically detect avalanches in airborne optical data. Lato et al. (2012) and Korzeniowska et al. (2017) applied object-based classification techniques to optical high-spatial-resolution data (0.25 - 0.5 m). Wesselink et al. (2017) and Eckerstorfer et al. (2019) have introduced and consequently refined an algorithm to automatically detect avalanches in Sentinel-1 SAR imagery, via changes in the backscatter between pre- and post-event images. Karbou et al. (2018) also utilized changes in backscatter to identify avalanche debris. For avalanche detection in RADARSAT-2 imagery, Hamar et al. (2016) used supervised classification with a random forest classifier. In contrast, the avalanche mapping from optical satellite data has so far been exclusively done manually (Bühler et al., 2019; Hafner et al., 2021; Abermann et al., 2019).

The deployment of machine learning for remote sensing image analysis has seen a surge in the last decade (Ma et al., 2019). Modern deep learning methods often outperform competing ones in complex image understanding tasks and have been used, for example, to detect rock glaciers (Robson et al., 2020), landslides (Prakash et al., 2021) and crop types in fields (Cai et al., 2018). For avalanches, the use of deep learning has so far focused on Sentinel-1 imagery: Waldeland et al. (2018) applied a pre-trained ResNet (He et al., 2016) for avalanche identification by change detection using manual reference annotations. Bianchi et al. (2021) segmented avalanches with a fully convolutional U-Net (Ronneberger et al., 2015), also relying on manual annotations for training the network. Sinha et al. (2019a) proposed a fully convolutional VGG16 network (Simonyan and Zisserman, 2015) that was trained on, and compared against, an inventory of avalanche field observations. With the same inventory, Sinha et al. (2019b) also, alternatively used a variational autoencoder (Kingma and Welling, 2019) for avalanche detection.

In contrast to previous studies, our work is the first to attempt to use deep learning for the detection of avalanches in *optical* satellite data. This is of major importance, as the largest avalanche mapping from remotely sensed imagery to date, with 24'778 single avalanche polygons (Bühler et al., 2019; Hafner and Bühler, 2019, 2021), relied on optical SPOT 6/7 satellite imagery. Furthermore, there have been investigations with external data into the reliability and completeness of mappings from SPOT 6/7 (Hafner et al., 2021). Consequently, an automation of the manual mapping from this imagery would allow for a fast comprehensive documentation of future avalanche periods with background knowledge about how well it works and how much avalanche area approximately is missed. Without an automation it is not feasible to cover large regions quickly. With manual image interpretation (Hafner et al., 2021) it took

approximately one hour to manually delineate avalanches in SPOT images covering a region of  $\approx 27.5 \text{ km}^2$ . Thus, in this work we develop, describe and apply a deep learning approach for avalanche mapping based on the SPOT 6/7 sensor with the goal to automate the mapping process, so as to cover large areas and eventually operate at country-scale. We developed a variant of DeepLabV3+ (Chen et al., 2018) that takes as input SPOT 6/7 images and a digital elevation model (DEM), and outputs spatially explicit raster maps of avalanches. For our DeepLabV3+ variant we made the encoder and decoder deformable (Dai et al., 2017), thereby our convolutional kernels adapt according to the underlying terrain, which is essential in the study of avalanches. In addition to a careful description of the network architecture we evaluate results, compare them to previous work, examine the reproducibility of the manually mapped avalanches, and discuss the potential and limitations of our method.

## Data

For training and validating our proposed mapping system we utilize SPOT 6/7 top-of-atmosphere reflectance images acquired on 24 January 2018 (referred to as 2018 in the remainder of this paper, Hafner and Bühler, 2019) and 16 January 2019 (referred to as 2019 from now on; Hafner and Bühler, 2021), together with a set of 24'776 avalanche annotations delineated by manual photo-interpretation. In both cases the images were acquired after periods with very high avalanche danger, i.e., the maximum level 5 of the Swiss avalanche warning system (WSL Institute for Snow and Avalanche Research SLF (ed.), 2021). SPOT 6/7 images have a ground sampling distance (GSD) of 1.5 m and provide information in four spectral bands, namely red, green, blue, and near-infrared (R, G, B, NIR), at a radiometric resolution of 12 bits. The dataset covers an area of  $\approx 12'500\text{km}^2$  in 2018 and  $\approx 9'500\text{km}^2$  in 2019. These two areas partly overlap. As both were acquired in January, the illumination conditions exhibit little variability between the two years, but they differ in terms of snow conditions: in 2019 the snow line was at a lower altitude, and consequently there was more dry snow, hardly any wet snow, and fewer glide snow avalanches. As additional input information we use the Swiss national DEM *swissALTI3D*. To match the resolution of SPOT imagery, we resample the DEM (original GSD 2 m) to 1.5 m, aligned with SPOT 6/7. Its nominal vertical accuracy is 0.5 m below the treeline ( $\sim 2100$  m a.s.l.) and 1–3 m above the treeline (swisstopo, 2018). We did not apply atmospheric corrections as our main focus is texture and the absolute spectral values do not matter for avalanche identification.

The 24'776 avalanches were annotated by a single person, an expert, whom we define as somebody very familiar with both satellite image interpretation and avalanches. For the mapping of avalanches the visual identification of crown and release areas, track and deposit through texture and hue, as well as hints of possible damage have played a role (for details on the methodology see Bühler et al., 2019). For each mapped avalanche polygon the expert also recorded a score of how well the avalanche was visible, splitting the annotations in three groups: *complete*, *well visible* outline; *mostly well visible* outline; and *not completely visible* outline, where significant parts had to be inferred with the help of domain knowledge (see also Bühler et al., 2019). Furthermore, we validated a subset of the initial mapping with independent ground- and helicopter-based photographs as reference (Hafner et al., 2021). We found that for manual mapping based on SPOT images the probability of detection (POD; see Equation 11.2; the probability of a true avalanche being annotated) is 0.74 for avalanches larger than size 1 (avalanche size is categorised on a scale from 1 to 5, with size 5 the largest and most destructive

ones; for more details see WSL Institute for Snow and Avalanche Research SLF (ed.), 2021). The positive predictive value (PPV; see Equation 11.2; probability of an annotated avalanche having a true counterpart) is 0.88, indicating only a few false positive annotations (again for size  $\geq 2$ ).

Additionally, we used SPOT 7 imagery of the Mattertal, Val d'Hérens and Val d'Herémence in Valais, Switzerland from 6 January 2018 covering  $\approx 660\text{km}^2$  to evaluate our model. The data were acquired for test purposes after a period with high avalanche danger, and the 538 avalanches used for validation have been manually mapped with the same methodology as the others used in this work and described in Bühler et al. (2019). The geographical region with additional data overlaps with data acquired on 24 January 2018, but served as test area before and did not go into training or validation (see "Generalization Test" areas in Figure 10.5). The images suffer from distortion in steep terrain as they were part of a suitability study for avalanche mapping from optical data (for details see Bühler et al., 2019) and orthorectified by the satellite providers using the height information from the Shuttle Radar Topography Mission (SRTM; OpenTopography, 2013).



## Method

Many overlapping avalanches exist in the dataset whose boundaries cannot be precisely distinguished from each other even by experts. We thus restrict ourselves to identifying all pixels where avalanches have occurred, but do not attempt to group them into individual avalanche events. In terms of image analysis this corresponds to a semantic segmentation task, where each pixel is assigned a class label: *avalanche* or *background*, according to the model confidence. Several deep learning models have been developed for solving such problems and have achieved excellent results in various domains, such as U-Net (Ronneberger et al., 2015), HRNetV2 (Sun et al., 2019) and DeepLabV3+ (Chen et al., 2018).

### 10.1 Model architecture

On their way downwards, avalanches are constrained and guided by the local terrain. In order to accurately map avalanches from the input data, we therefore propose a deep learning architecture that adapts to the underlying terrain model. We build on the state-of-the-art model DeepLabV3+ designed for semantic segmentation and add deformable convolutions that adapt their receptive field size according to the input data, i.e., the terrain model in our case.

**DeepLabV3+** is a popular, fully convolutional semantic segmentation model that has been used successfully with a variety of datasets. It features a dilated ResNet (He et al., 2016) encoder as a backbone for feature extraction, in combination with the Atrous spatial pyramid pooling module (ASPP). To achieve a wide receptive field able to capture multi-scale context, ASPP employs dilated convolutions at different rates. Before being fed into the decoder, the resulting features are concatenated and merged using a  $1 \times 1$  convolution. These high-level features are then decoded, upsampled and combined with high-resolution, low-level features from the first encoder layer. For further details about DeepLabV3+, see Chen et al. (2018).

Our adaptations to the standard DeepLabV3+ include *deformable kernels* (Dai et al., 2017) in the encoder and decoder, as well as a small network with offsets that estimates the appropriate kernel deformations in a data-driven manner and modifies the decoder such that it can process features from *all backbone layers* (Figure 10.1 and Figure 10.2). These changes add a modest 1.9 million network weights to the 22.4 million weights of the standard DeepLabV3+.

The reasoning behind *deformable convolution kernels in the backbone* (Figure 10.3) is to adapt their receptive fields to the underlying terrain. To obtain deformable convolutions, we introduce an

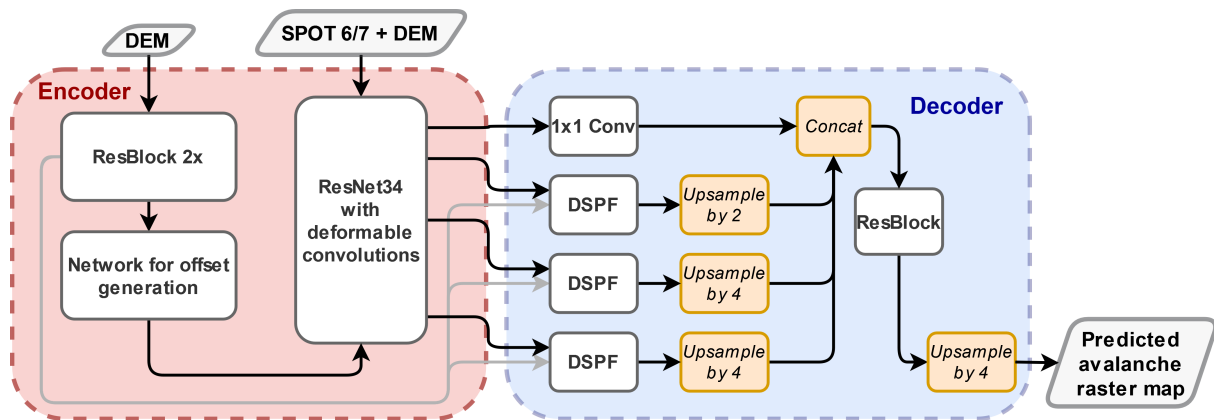


Figure 10.1: Overview of our DeepLabV3+ variant. The encoder is shown in more detail in Figure 10.3 and the deformable spatial pyramid flow (DSPF) in Figure 10.4.

additional 18-channel tensor that encodes the 2D offset of each kernel element at each location i.e., it enables free-form deformations of the kernel, beyond dilation or rotation. The offsets are not fixed a priori but calculated as a learned function of the DEM, separately for each feature resolution, by a small additional network branch. By replacing the first convolution in each residual block with a deformable one, we are able to explicitly include the terrain shape encoded in the DEM, but without the need to modify other parts of the architecture, so as to benefit from the pretrained weights of the encoder.

The *augmented decoder* helps our DeepLabV3+ to propagate features along specific directions, in our case this is the possible downhill flow direction of avalanches which can be extracted from the DEM. Hence, we alter the ASPP such that it aggregates features from *all* backbone layers, and increases the receptive field. The new module, which we call *deformable spatial pyramid flow* (DSPF, Figure 10.4), performs deformable convolutions at different dilation rates. The deformations are again obtained from our small network with offsets, based on the DEM. In order to propagate information along the gradient field, we also model the flow direction of an avalanche in the DSPF module of the decoder.

### 10.1.1 Sampling and data split

Given the proposed model architecture and the available computational resources (CPU: 20 Intel Core 3.70 GHz, GPU: 1 NVIDIA GeForce RTX 2080 Ti), we are unable to process an entire

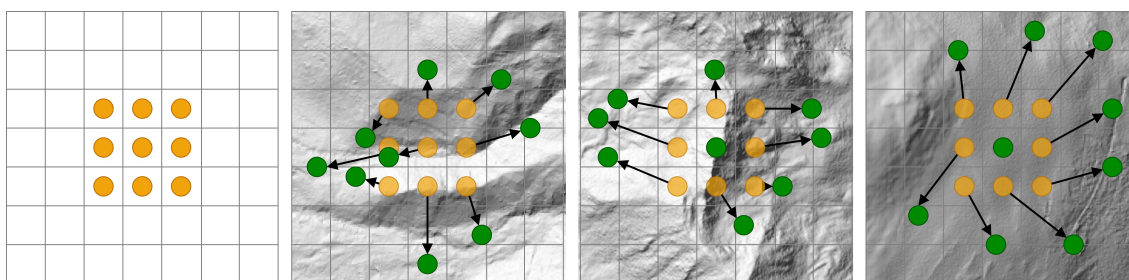


Figure 10.2: For the deformable convolutions, a standard kernel (like the  $3 \times 3$  as shown in a) will be adapted according to 2D offsets learned from the underlying DEM. The green dots in (b), (c) and (d) exemplarily show possible final positions of the kernel elements, the displacement from the standard kernel is illustrated by the black arrows.

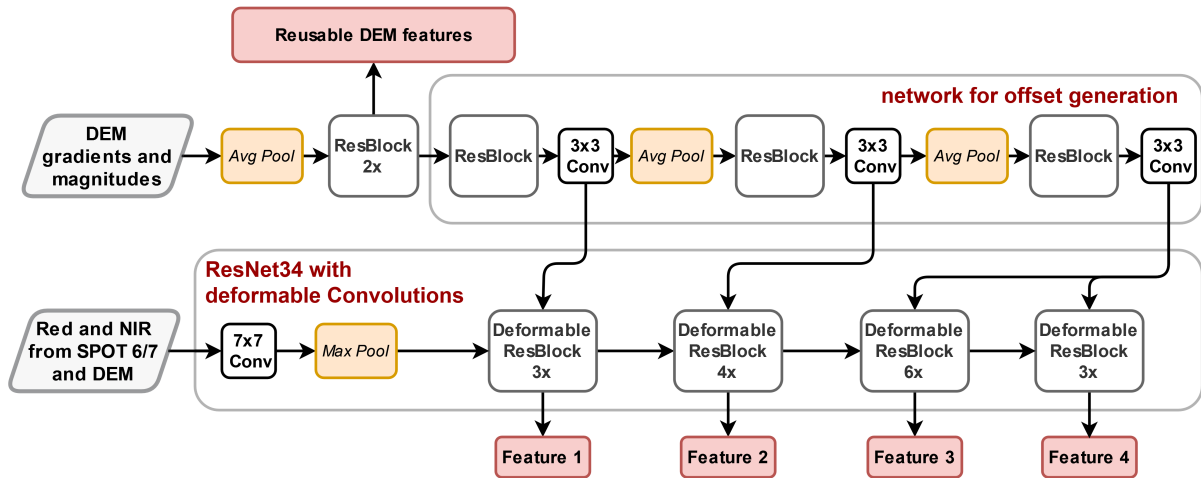


Figure 10.3: Encoder of our DeepLabV3+ in detail.

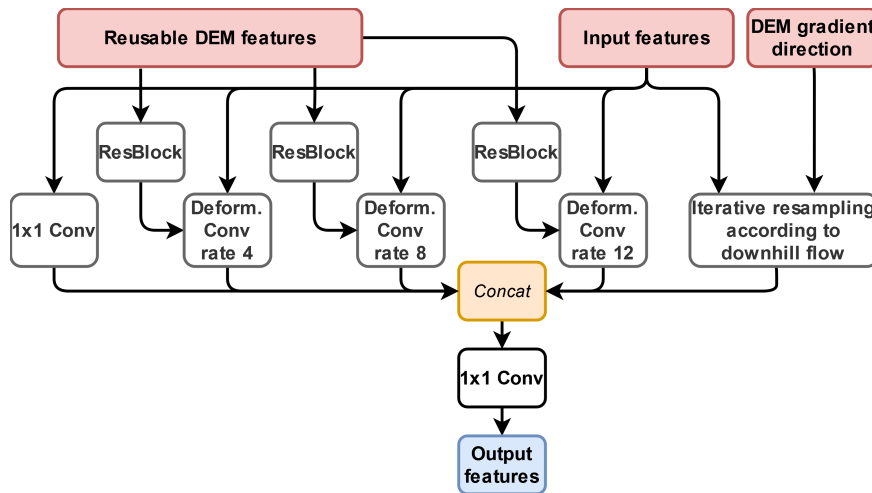


Figure 10.4: Detailed architecture of the deformable spatial pyramid flow (DSPF) used in the decoder of our DeepLabV3+ variant.

orthomosaic at once. Therefore, we process squared image subsets, called patches, of up to  $512 \times 512$  pixels at training time, which translates into an area of  $589'824 \text{ m}^2$  at the spatial resolution of SPOT 6/7 imagery. With our model and computational resources we can simultaneously process batches of two image patches per GPU.

For supervised machine learning approaches it is vitally important that all desired classes are present in the patches the model learns from. As classes are usually not evenly distributed, class imbalance is a frequent challenge. Our dataset is very imbalanced: avalanches cover only  $1/1785$  of the entire area covered by SPOT 6/7 imagery. Re-balancing of class frequencies is necessary to make sure our model adequately captures the variability of the avalanche class. We use the following pragmatic strategy to ensure a training set that includes relevant examples and with sufficient representation of both classes. First, we iteratively sample patch centers inside manually annotated avalanche polygons while avoiding overlapping patches. In this way, we obtain a set of samples that is not overly imbalanced, with  $\approx 3.5 \times$  more background pixels than avalanche pixels. These patches form 95% of our training set. Second, the remaining 5% are

sampled randomly in areas without avalanches to ensure also patches without avalanche pixels are seen during training. This leads to an effective ratio of 1:4 between avalanche and background pixels in the 5185  $512 \times 512$  patches of the training set.

As the edges of the patches lack context, they were also given smaller weights when calculating the loss function during training, starting 100 pixels from the edge, decreasing the weight linearly to 10% of the base weight given above at the very edge. For our DeepLabV3+ we additionally used deep supervision as in Simonyan and Zisserman (2015) to help the model converge.

### 10.1.2 Training

For training and quantitative evaluation, the data were split into mutually exclusive, geographically disjoint regions for training (80%), validation and hyper-parameter tuning (10%) and testing (10%), as depicted in Figure 10.5. The test set is located completely in regions acquired either only in 2018 or only in 2019 but not in the overlap between the two acquisitions to prevent memorization (especially of the identical topography).

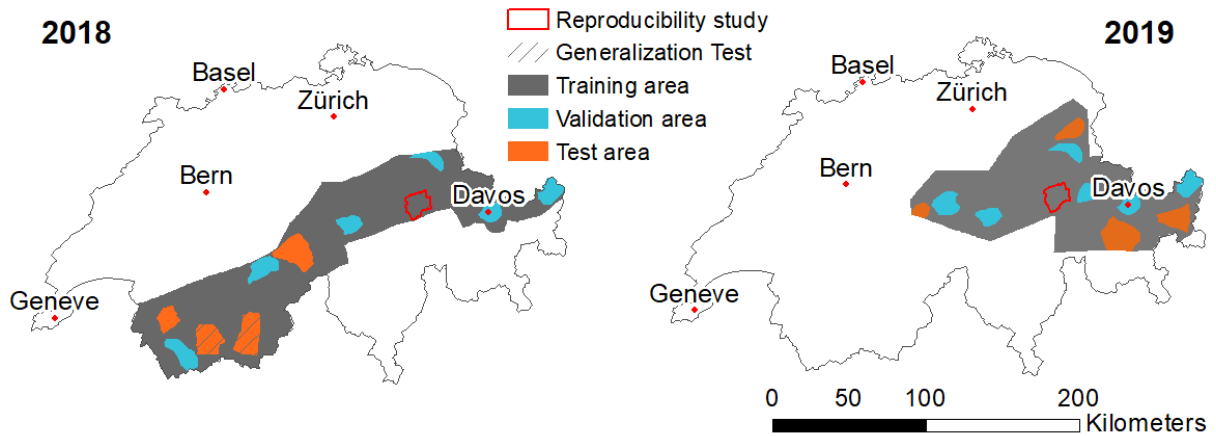


Figure 10.5: Visualization of the disjoint regions for training, validation and testing for both 2018 and 2019. Also shown are the test region for the generalization experiments, where we had additional data from 6 January 2018, and the regions used to study reproducibility of manual avalanche maps.

The network is trained by minimizing a weighted binary cross entropy (BCE) loss (see also 10.1.2), using the Adam optimizer (Kingma and Ba, 2017) for 20 epochs. The base learning rate was initialized to  $1 \times 10^{-4}$  and reduced by a factor of 4 after 10 epochs. A summary of the hyper-parameter settings is given in Table 10.1.

Table 10.1: Summary of training parameters

Parameter	Value
Loss function	Weighted BCE
Optimizer	Adam
Initial learning rate	$1 \times 10^{-4}$
Effective batch size	16
Patch size	$512 \times 512$
Epochs	20

As a preprocessing step, the input images are normalized channel-wise using the mean and

variance values of the entire dataset. Additionally, we flattened the peak in the image histograms caused by the shadow pixels by transforming negative values  $v \rightarrow (-3 \cdot v^2)$  while keeping positive values unchanged.

Even though our training dataset is large, it covers only two avalanche periods and cannot be expected to account for the whole variety of possible conditions. In order to increase the robustness of the network, we further expand the training set with synthetic data augmentation. We used randomized rotation and flipping for greater topographic variety, mean-shifting and variance-scaling to simulate varying atmosphere and lighting conditions, and patch shifting to increase robustness when only part of an avalanche is visible. To speed up data loading we used batch augmentation (Hoffer et al., 2019), in which the same sample is read only once and used multiple times with different augmentations computed on the fly. To increase the model's performance, we additionally accumulated gradients over two iterations before weights were updated. Thereby an effective batch size of four (2+2) was reached, and the  $512 \times 512$  pixel patches may be used (see also Section 11.2).

As mentioned in Section 9 the avalanche polygons come with labels that quantify their visibility in the SPOT data. These labels are used to re-weight their contributions to the BCE loss as follows: pixels on *complete, well visible* avalanches have weight 2, *mostly well visible* avalanches, as well as *background* pixels not on an avalanche, have weight 1, and *not completely visible* avalanches have weight 0.5.

## Results and discussion

Predictions are made for a target area specified by vector polygons in the form of shapefiles. To reduce artifacts at the edges of patches, the samples for the predictions overlap by 100 pixels before being cropped. To assess the detection performance of the network, we calculated positive predictive value (PPV, also called precision) and probability of detection (POD, also called recall) on a pixel level, as well as the F1 score. PPV and POD are both based on a standard  $2 \times 2$  confusion matrix (Trevethan, 2017). As per-pixel metrics take as input a binary mask (avalanche yes or no) and the network yields scores, we thresholded the predictions at 0.5 before calculating statistics and computed the F1 score as

$$F1 = 2 \cdot \frac{PPV \cdot POD}{PPV + POD}, \quad (11.1)$$

where POD and PPV are defined as

$$POD = \frac{TP}{TP + FN} \quad \text{and} \quad PPV = \frac{TP}{TP + FP}, \quad (11.2)$$

where TP is true positive, FP is false positive and FN is false negative.

In this paper the presented pixel-wise metrics (POD, PPV and F1 score) represent the average score over all the patches we tested on. As our dataset is imbalanced and the F1 score non-symmetric, we calculated those metrics for both avalanches and the background. Additionally, we wanted to estimate how many avalanches were detected by each model. Consequently, for the object-based metrics we tested two different measures: we counted an avalanche as detected if 50% or 80% of all pixels within an avalanche from the manual mapping had a score of 0.5 or higher.

### 11.1 Results and generalization ability

Results were calculated for the test areas and are reported in Table 11.1. Compared to the standard DeepLabV3+, our model, when run with the parameters described in Table 10.1, has a higher POD for avalanches (0.610 vs. 0.587) while having the same PPV. This results in an F1 score of 0.612 for the standard DeepLabV3+ and 0.625 for our version. For the background, the pattern is similar: the POD is slightly better for our method (0.894), compared to the standard DeepLabV3+ (0.888), while the PPV is slightly higher for the standard model (0.900 vs. 0.894).

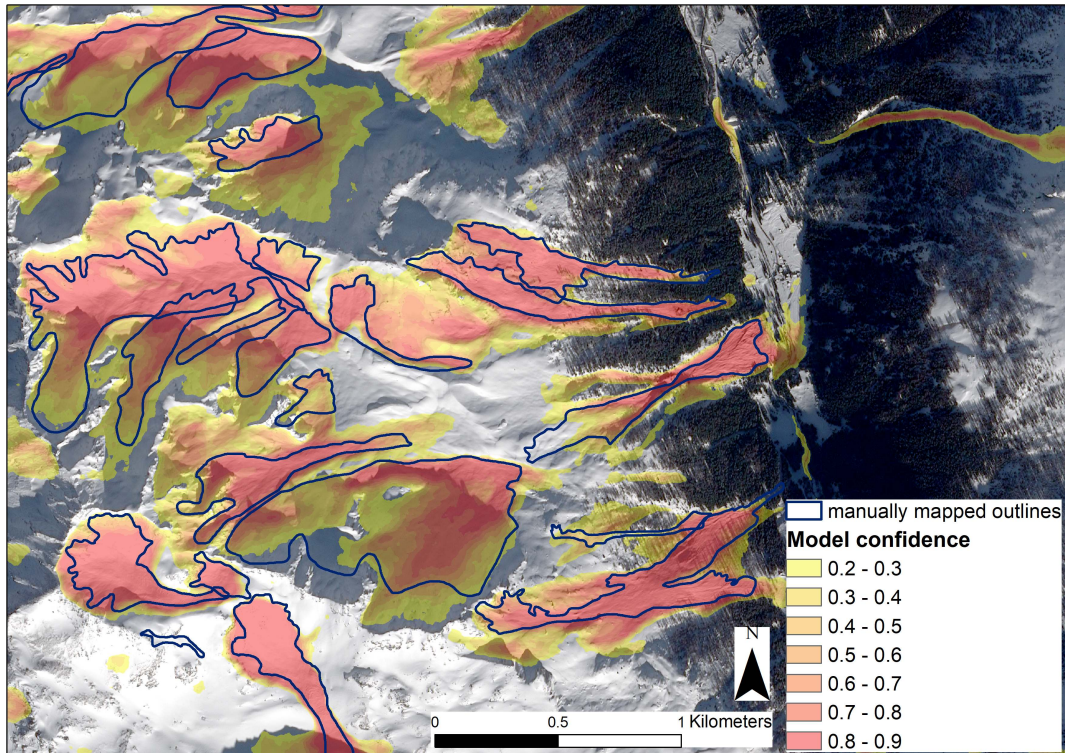


Figure 11.1: An example for the model confidence when predicting on data from a previously unseen avalanche period from 6 January 2018 (SPOT 6 data © Airbus DS 2018). The values closer to 1, in darker hues, indicate places where the model is more confident about the existence of an avalanche. In the illuminated regions those areas almost always overlap with manually mapped avalanches.

Consequently, the F1 score is very similar as it only differs by one in the third decimal place between our and the standard DeepLabV3+.

For any supervised classification and deep learning methods in particular, the ability to generalize well to new datasets and regions not seen during the training phase is key. To evaluate this, we test our trained model using SPOT 7 imagery from 6 January 2018. The test metrics for predictions on the data from 6 January 2018 were calculated with the standard DeepLabV3+ and the adapted DeepLabV3+. As Table 11.1 shows, our version generalizes very well (see also Figure 11.1), the metrics only differ from tests on the initial dataset in the fourth decimal place. The standard DeepLabV3+ on the other hand, does not generalize so well as the POD and the detection rates per avalanche are lower than for testing on the initial data.

We also investigated object-based metrics for all model variations, when detection means 50% of the avalanche area the models rightly capture between roughly 58% and 69% of all avalanches and between 38% and 51% when detection requires 80% of the area (Table 11.2). Again the the standard DeepLabV3+ performs slightly worse than our adapted DeepLabV3+, especially when run on data from a new avalanche period (6 January 2018). Therefore, our DeeplabV3+ shows better ability to generalize to new and previously unseen data. Overall, the best performance is achieved when considering sunlit avalanche parts only, for both training and testing.

Table 11.1: Segmentation results for the test areas for the standard DeepLabV3+ and our DeepLabV3+, results of predicting on data from a previously unseen avalanche period and variations to our model for the ablation. The metrics shown are the averages from all tested patches. The bold fonts signify the model, data and parameters that were varied compared to what we used as standard in our work.

Model	SPOT data	Training data	Patch size	Bands	Loss function	POD	PPV	F1	POD back	PPV back	F1 back
<b>Standard</b>	24.01.2018	Whole	512×512	R, NIR, DEM	Weighted	0.587	0.667	0.612	0.888	0.900	0.885
<b>DeepLabV3+</b>	16.01.2019	avalanches			BCE						
Adapted	24.01.2018	Whole	512×512	R, NIR, DEM	Weighted	0.610	0.668	0.625	0.894	0.894	0.884
DeepLabV3+	16.01.2019	avalanches			BCE						
<b>Standard</b>	<b>06.01.2018</b>	Whole	512×512	R, NIR, DEM	Weighted	0.547	0.667	0.591	0.876	0.916	0.887
<b>DeepLabV3+</b>		avalanches			BCE						
Adapted	<b>06.01.2018</b>	Whole	512×512	R, NIR, DEM	Weighted	0.610	0.668	0.625	0.894	0.895	0.884
DeepLabV3+		avalanches			BCE						
Adapted	24.01.2018	Whole	<b>256×256</b>	R, NIR, DEM	Weighted	0.723	0.587	0.645	0.720	0.796	0.720
DeepLabV3+	16.01.2019	avalanches			BCE						
Adapted	24.01.2018	Whole	<b>128×128</b>	R, NIR, DEM	Weighted	0.898	0.659	0.829	0.340	0.551	0.452
DeepLabV3+	16.01.2019	avalanches			BCE						
Adapted	24.01.2018	Whole	512×512	R, NIR, DEM	<b>unWeighted</b>	0.688	0.575	0.622	0.887	0.908	0.888
DeepLabV3+	16.01.2019	avalanches			<b>BCE</b>						
Adapted	24.01.2018	Whole	512×512	<b>R, G, B, NIR, DEM</b>	Weighted	0.559	0.682	0.613	0.883	0.913	0.889
DeepLabV3+	16.01.2019	avalanches		<b>DEM</b>	BCE						
Adapted	24.01.2018	Whole	512×512	<b>R, G, B, NIR, DEM, Wallis-filtered</b>	Weighted	0.610	0.668	0.597	0.876	0.906	0.880
DeepLabV3+	16.01.2019	avalanches			BCE						
Adapted	24.01.2018	<b>Release area only</b>	512×512	R, NIR, DEM	Weighted	0.053	0.665	0.183	0.777	0.992	0.856
DeepLabV3+	16.01.2019	<b>area only</b>			BCE						
Adapted	24.01.2018	<b>Deposits only</b>	512×512	R, NIR, DEM	Weighted	0.196	0.788	0.347	0.797	0.986	0.868
DeepLabV3+	16.01.2019	<b>only</b>			BCE						
Adapted	24.01.2018	<b>Sunlit avalanche parts only</b>	512×512	R, NIR, DEM	Weighted	0.668	0.653	0.639	0.918	0.910	0.907
DeepLabV3+	16.01.2019	<b>avalanche parts only</b>			BCE						



Table 11.2: Object-based metrics for selected model configurations.

Model	SPOT data	Training data	Detection rate 50% of avalanche area	Detection rate 80% of avalanche area
Standard DeepLabV3+	24.01.2018 16.01.2019	Whole avalanches	0.63	0.45
Adapted DeepLabV3+	24.01.2018 16.01.2019	Whole avalanches	0.66	0.46
Standard DeepLabV3+	06.01.2018	Whole avalanches	0.58	0.38
Adapted DeepLabV3+	06.01.2018	Whole avalanches	0.66	0.46
Adapted DeepLabV3+	24.01.2018 16.01.2019	Sunlit avalanches only	0.69	0.51

## 11.2 Ablation studies

To understand how our changes to the standard DeepLabV3+ affect performance we varied the model in different ways and trained, tested and compared the performance. These results can be found in Table 11.1. First, we investigated the influence of the deformable backbone and discovered that including it outperforms the non-deformable backbone configurations of the standard DeepLabV3+. This is the case in our test areas for 2018 and 2019, but also for testing on the avalanche period from 6 January 2018. Secondly, the avalanches in our network have been weighted (see 10.1.2) according to the quality index assigned by the manual mapper. To quantify the effects of using weights we ran training with unweighted BCE and observed a decrease in POD, a slight increase in PPV and overall a smaller F1 score. Additionally, in our adapted version of DeepLabV3+ we only considered the red and near-infrared from SPOT, as well as the DEM as input channels. We cannot test the adapted DeepLabV3+ without the DEM as it is explicitly included as an integral part of the network. We analyzed, however, how including all SPOT channels (additionally Blue and Green) and also adding another Wallis filtered channel (to bring out details in the shade) affect network performance (see Table 11.1). For our model we found that including more channels did not improve the performance; rather training time was longer and metrics worse than with the initial channels.

We hypothesize that the proportion of potential avalanche area and context visible in the patches strongly influences network output. To investigate this, we have trained our model with varying patch sizes:  $512 \times 512$ ,  $256 \times 256$  and  $128 \times 128$  pixels (corresponding to  $768 \times 768$ ,  $384 \times 384$  and  $192 \times 192$  meters). Quantitative results in Table 11.1 show the largest patch size performs best considering metrics for both avalanches and background. When comparing them visually (Figure 11.2), this is further supported as the predictions on the smallest size are patchy and dispersed over the image, showing the model is unsure about the occurrence of avalanches. With increasing context through a larger patch size though, the model becomes more confident, and the avalanche borders are distinctly visible.

Subsequently, in order to understand what is better for training the network, we trained on avalanche deposits or release areas only. As deposit area, we assumed the lower third (based on elevation) of each manually mapped avalanche, ignoring those avalanches where the deposit had been inferred. For the release areas, we used the zones identified by Bühler et al. (2019), again

disregarding those avalanches where the release zone had been inferred and was therefore uncertain. As results in Table 11.1 show, performance for predicting all avalanches is a lot worse in both cases. We also observe that PPV and POD are significantly higher when the network is trained on deposits only rather than trained only on release areas. This resulted in an increase of 0.146 in F1 score and suggests that the original model might also be learning more from texture-rich avalanche deposits than from release zones.

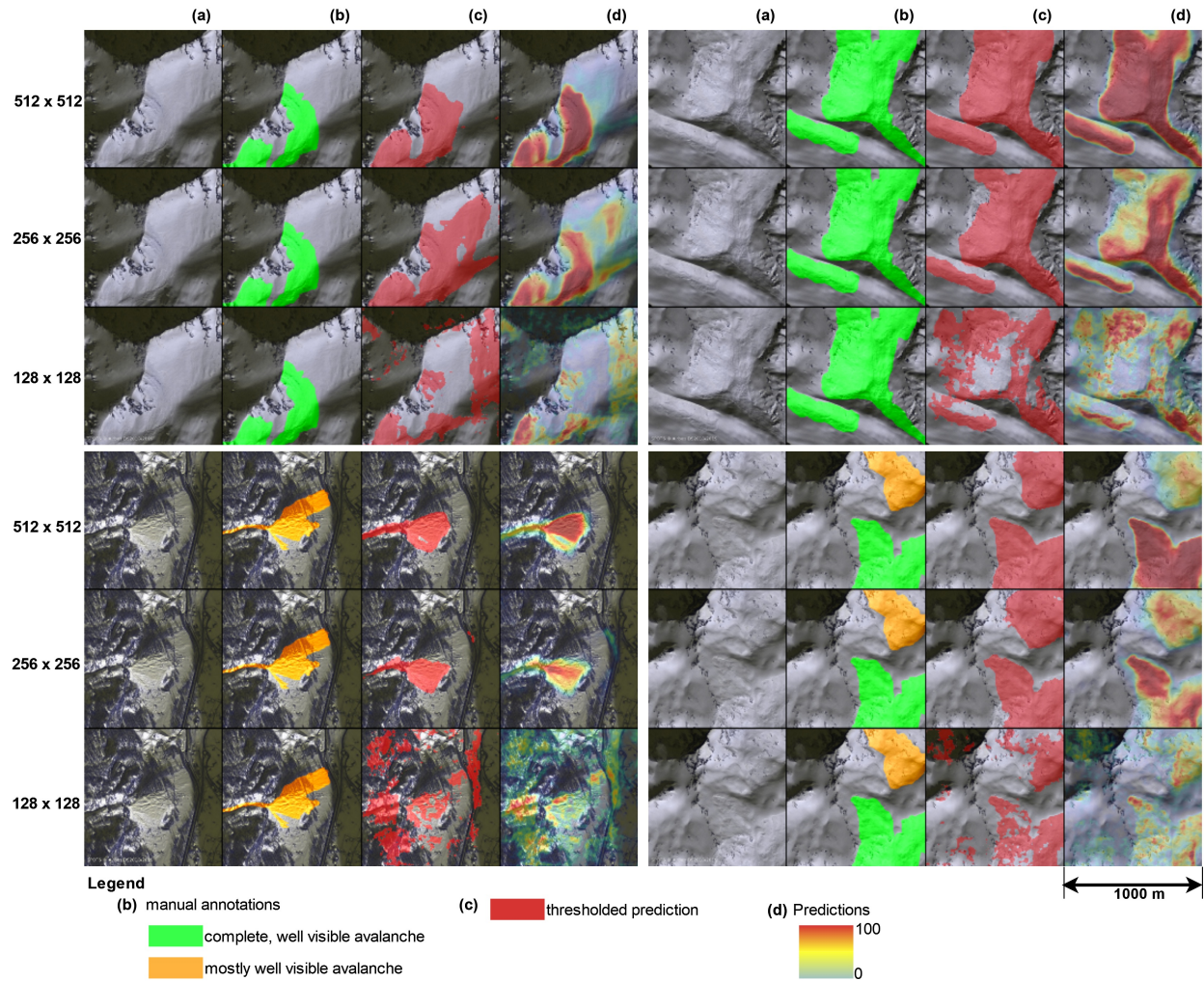


Figure 11.2: Comparison of results for four patches when training the network with different patch sizes. The tiles depict a) the SPOT 6 image, b) the manually mapped annotations used as reference, c) the predictions thresholded at 0.5 and d) the predicted avalanche probability (SPOT 6 data © Airbus DS 2018). Visual inspections show that the model is a lot more confident the larger the patch size is.

Finally, the experts manually mapping the avalanches generally perceived those in the sun as better visible. Hafner et al. (2021) confirmed that and found the POD to be higher roughly by a factor of 5 for avalanches in fully illuminated terrain compared to those, at the time of image acquisition, in fully shaded terrain. In order to investigate this further, we used a support vector machine (SVM) classifier to calculate a shadow mask for both 2018 and 2019. The mask also includes most forested areas due to their speckled sun-shade pattern. Subsequently, we excluded the avalanche parts located in the shade and trained only with the remaining areas

(about one fourth of the avalanche area per year). Calculating the metrics considering only avalanches in illuminated areas, we found an increase of 0.058 in POD, a slight decrease of 0.015 in PPV and consequently an increase in F1 score of 0.014. The object-based metrics (Table 11.2) are also slightly better when only considering sunlit regions.

### 11.3 Reproducibility of manually mapped avalanches

To assess the degree of label noise in our dataset, we conducted a reproducibility experiment on the manually mapped avalanches to understand how similar the assessment of a given area by different experts would be. In other fields several comprehensive studies have already been conducted to investigate inter-observer variability, for example for contouring organs in medical images (Fiorino et al., 1998) or for manual glacier outline identification (Paul et al., 2013). For our investigation five people attempted to replicate the manual mapping with the same methodology as used before and described in detail in Bühler et al. (2019). All five mapping experts are very familiar with satellite imagery and/or avalanches and received the same standardized introductions. The experiment was conducted twice in an area of 9km<sup>2</sup> around Flims, Switzerland, on the 2018 and 2019 SPOT 6/7 imagery (see Figure 10.5). The area contains avalanches in the shade and in illuminated terrain as well as all outline quality classes in the initial mappings (Hafner and Bühler, 2019, 2021). The mapping experts did not see another mapping before having finished theirs.

Table 11.3: F1 scores for the reproducibility investigation: the bold values in the upper-right part of the table represent the scores comparing two expert mappings in illuminated terrain, and the lower-left values are the scores in shaded terrain.

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Expert 1		<b>0.758</b>	<b>0.623</b>	<b>0.617</b>	<b>0.653</b>
Expert 2	0.401		<b>0.711</b>	<b>0.723</b>	<b>0.724</b>
Expert 3	0.232	0.198		<b>0.656</b>	<b>0.782</b>
Expert 4	0.188	0.236	0.205		<b>0.786</b>
Expert 5	0.123	0.155	0.204	0.244	

Calculating F1 score (see Equation 21.1) between all experiment mappings, we found an overall F1 score of 0.381 in illuminated and 0.018 in shaded areas (area-wise metrics). Comparing two expert mappings at a time, the values range from 0.617 to 0.786 in the illuminated regions and from 0.123 to 0.401 in the shaded regions of our study area (Table 11.3). The F1 scores of the expert manual mappings with the initial mapping are in the same range (not shown). The results from 2018 (Figure 11.3) illustrate that for some selected avalanches the agreement is very good, while, especially in the shade, there is little agreement among experts on the presence of avalanches.

Reexamining the results from the network now in the light of this experiment, the adapted DeepLabV3+ is equally good as the experts in identifying avalanches. In other words, we cannot expect a computer algorithm to provide better scores than the average F1 score of two mapping experts. Even for the avalanches with the highest agreement, a specific boundary line will usually not match exactly. This makes it hard for any network to learn the localization of boundaries. We do not yet know what exactly causes the differences in avalanche identification between experts. Therefore we plan on conducting a thorough analysis on imagery with different spatial resolutions in the future. This will help us to better understand the inherent mapping uncertainty of avalanches and may give an indication of what performance can be expected if training computational detection algorithms on different optical data.

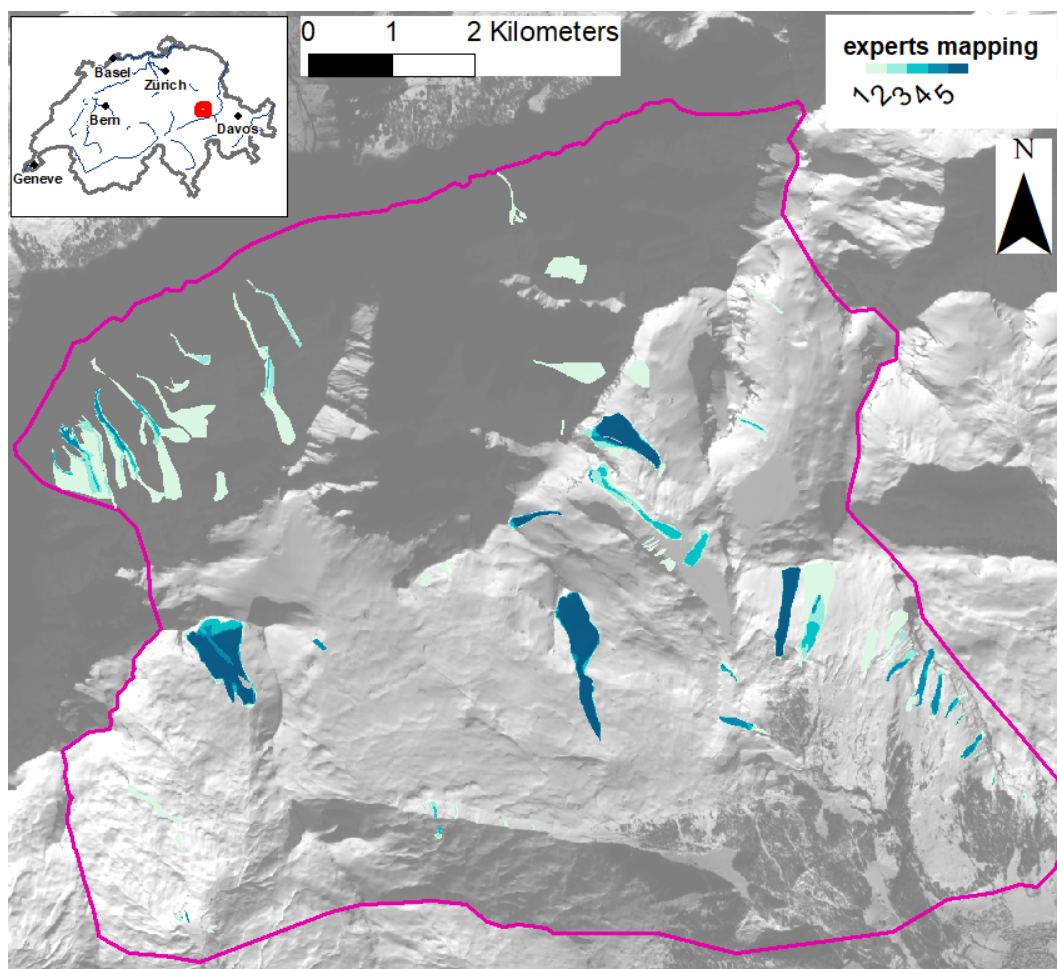


Figure 11.3: Heat map exemplarily illustrating expert agreement on avalanche area for avalanches mapped from SPOT in January 2018 (24 January 2018, SPOT 6 © Airbus DS2018). Agreement in the shade (northern part of the study area) is generally lower than in the sunlit areas to the south. Dark blue indicates very good agreement or in other words marks areas that were identified as an avalanche by all five experts involved. For a more detailed location of the reproducibility study area see Figure 10.5.

## 11.4 Limitations of this study

The three avalanche periods for which we have SPOT imagery all occurred in January. Those images are relatively close to the winter solstice and therefore have a high percentage of shaded area. The amount of shaded area depends very much on the terrain and on the season. Around Davos, Switzerland, for example, 43% of the area is shaded on the winter solstice but only 7% 3 months later (both at SPOT 6/7 image acquisition time; Hafner et al., 2021). We know that the quality of the manually annotated avalanches is lower in shaded areas (POD: 0.15 shade, 0.86 illuminated, 0.74 overall; Hafner et al., 2021). Consequently, the training data have lower quality in shaded regions, which makes learning there more difficult for our model and leads to lower model confidence, as well as poorer results. Based on the results when training and testing on sunlit avalanche parts only, however, we see potential for better overall metrics when a smaller portion of the area is shaded closer to the summer solstice. But regardless of how much area is well illuminated, the challenges in the shade remain and make results in those areas less

trustworthy. Further research to better understand and tackle that problem is needed.

Additionally, even though 2018 includes wet snow and wet snow avalanches, the snow in January is generally colder and drier than towards the end of the winter. Consequently, we do not know how well our model performs under different snow conditions, for example in spring. Whether our model already generalizes enough or is biased towards high winter conditions and requires retraining with different snow conditions, we could not yet test.

## Conclusion and outlook

We present a novel deep learning approach for avalanche mapping with deformable convolutions that adapts its notion of the local terrain according to the input digital elevation model (DEM). Experiments at large scale with optical, high-spatial-resolution (1.5 m) SPOT 6/7 satellite imagery show that our approach achieves good performance (F1 score 0.625) and generalizes well to new scenes not seen during the training phase (F1 score 0.625). As reference data for training, validating and testing our model we relied on 24'747 manually mapped and annotated avalanches from two avalanche periods in different years. With our adapted DeepLabV3+ we were able to detect 66% of all avalanches. By varying model parameters and the input data we analyzed the impact of different configurations on the mapping result. We found that weighting the avalanches according to the perceived visibility did result in slightly better metrics than when not weighting them. By training on release areas and deposits only we demonstrated that the network learns more from deposits (Table 11.1), and by excluding shaded areas from training we showed that in illuminated terrain both training is easier and test results are better (F1 score 0.639). Furthermore, we investigated expert agreement for manual avalanche mapping in a small reproducibility study and found that agreement on avalanche area is substantially lower than expected. Compared to the model, the agreement between experts is in the same range as the adapted DeepLabV3+ performance.

Our work is an important step towards a fast and comprehensive documentation of avalanche periods from optical satellite imagery. This could substantially complement existing avalanche databases, improving their reliability to perform hazard zoning or the planning of mitigation measures. For the future we aim at conducting a more through study investigating expert agreement for manual avalanche identification and its implications for automated avalanche mapping. Additionally, we intend to study the performance of our model on data from different sensors and time periods. Furthermore, we plan on improving results by masking out areas where avalanches cannot occur using, for example, modeled avalanche hazard indication data from Bühler et al. (2022).

# Part III

## Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice

Elisabeth D. Hafner, Frank Techel, Rodrigo Caye Daudt, Jan Dirk  
Wegner, Konrad Schindler, Yves Bühler

**Natural Hazards and Earth System Sciences, published: 30.08.2023**

**Author version:** for the published typeset version please refer to Hafner et al. (2023b)

**Author Contributions:** EDH designed the size survey with input from FT, initiated and coordinated the avalanche mappings, performed the analyses in close collaboration with the co-authors, and wrote the paper draft. FT and EDH were part of the team of experts mapping from photographs. EDH and YB were part of the team of experts mapping from remotely sensed imagery. FT delivered the necessary input from the avalanche warning service, critically reviewed the results, and heavily contributed to the analysis and writing of the paper. All co-authors reviewed and complemented the manuscript.

## **Abstract**

Consistent estimates of avalanche size are crucial for communicating not only among avalanche practitioners but also between avalanche forecasters and the public, for instance in public avalanche forecasts. Moreover, applications such as risk management and numerical avalanche simulations rely on accurately mapped outlines of past avalanche events. Since there is not a widely applicable and objective way to measure avalanche size or to determine the outlines of an avalanche, we need to rely on human estimations. Therefore, knowing about the reliability of avalanche size estimates and avalanche outlines is essential as errors will impact applications relying on this kind of data. In the first of three user studies, we investigate the reliability in avalanche size estimates by comparing estimates for 10 avalanches made by 170 avalanche professionals working in Europe or North America. In the other two studies, completed with two pilot studies, we explore reliability in the mappings of six avalanches from oblique photographs from 10 participants and the mappings of avalanches visible on 2.9 km<sup>2</sup> of remotely sensed imagery in four different spatial resolutions from 5 participants. We observed an average agreement of 66% in the most frequently given avalanche size, while agreement with the avalanche size considered «correct» was 74%. Moreover, European avalanche practitioners rated avalanches significantly larger for 8 out of 10 avalanches, compared to North Americans. Assuming that participants are equally competent in the estimation of avalanche size, we calculated a score describing the factor required to obtain the observed agreement rate between any two size estimates. This factor was 0.72 in our dataset. It can be regarded as the certainty related to a size estimate by an individual, and thus provides an indication of the reliability of a label. For the outlines mapped from oblique photographs, we noted a mean overlapping proportion of 52% for any two avalanche mappings and 60% compared to a reference mapping. The outlines mapped from remotely sensed imagery had a mean overlapping proportion of 46% (image resolution of 2 m) to 68% (25 cm) between any two mappings and 64% (2 m) to 80% (25 cm) when compared to the reference. The presented findings demonstrate that the reliability of size estimates and of mapped avalanche outlines is limited. As these data are often used as reference data or even ground truth to validate further applications, the identified limitations and uncertainties may influence results and should be considered.



## Introduction

Information on the location and size of avalanches is crucial for avalanche forecasting. A consistent and accurate documentation of the outlines of avalanches is important for applications such as avalanche-related risk management, hazard mitigation measures or hazard zone planning. In addition, these data are used as ground truth, for instance, for the validation of numerical avalanche simulations (e.g., Wever et al., 2018), when training models for the automated detection of avalanches on satellite images (e.g., Hafner et al., 2022) or for training models estimating avalanche size from snowpack simulations (e.g., Mayer et al., 2023a). However, avalanche size estimates are subjective as they cannot easily be measured like, for instance, earthquakes. The same is true for avalanche outlines, where no objective way of determining them exists. In many applications where direct measurements are not possible, human estimates are used as the reference, sometimes referred to as a «gold standard» (e.g., Weller and Mann, 1997). Applications, where such data are used, include mapping of landslides (Ardizzone et al., 2002; Galli et al., 2008), identifying rock glaciers (Brardinoni et al., 2019), or the estimation of avalanche size and danger (e.g., Schweizer et al., 2020). When these data are used for validation, errors in the estimates may cause an observed reduction in model or forecast performance, simply due to errors in the reference (e.g., Bowler, 2006; Lampert et al., 2016). Therefore, quantifying reliability, defined as the consistency of repeated measurements or judgments of the same event relying on the same process (Cronbach, 1947), is vital.

The reliability of judgments of something that cannot be known directly may be described using Brunswik's lens model (Stewart, 2001): the parameter that cannot be directly measured is estimated using the information available (data). These data may be imperfectly describing the parameter of interest. The connection between the parameter and the actual event is the accuracy of the estimate. It may be reduced by unreliability either in the information (data) or in the information processing for making the judgment (skill of the judge; Stewart, 2001). The reliability of judgments may be approximated with repeated estimates, regression models or the measurement of agreement among estimates (Stewart, 2001). Such investigations, for comparable tasks where human estimates are important, have revealed that the automated mapping of clean glacier ice is at least as accurate as manual digitization (Paul et al., 2013). Galli et al. (2008) found the time available for field reconnaissance to correlate with the accuracy of landslide event inventory maps. Brardinoni et al. (2019) analyzed the variability in rock glacier inventories and found it to depend, in comparable proportions, on inter-operator variability and

the quality of available imagery.

Since both avalanche size and avalanche outlines are currently assessed relying on human interpreters, and since consistent and accurate size estimates and avalanche outlines are key data for several applications, our objective is to quantify the reliability of these data. We expand previous studies exploring the consistency in avalanche size estimates (Moner et al., 2013; Jamieson et al., 2014) using a larger sample. Moreover, we quantify the reliability in avalanche outlines mapped from oblique photographs and remotely sensed imagery, and investigate potential factors explaining inter- and intra-rater variations.

In three independent user studies we address the following two research questions:

1. To what degree do experts agree when rating the size of an avalanche from photographs?
2. To what degree do experts agree when mapping the outline of avalanches from oblique photographs or from remotely sensed imagery?

Moreover, we explore potential factors influencing the agreement rates in size estimates or avalanche outline mappings. This allows for the estimation of benchmark values describing the reliability of these kind of data, and hence the interpretation of the performance of applications relying on these data.

## Background

Avalanche size may be assessed by installing infrastructure to measure impact pressure (e.g., Sovilla et al., 2008) or by determining deposit volumes with photogrammetry (e.g., Eckerstorfer et al., 2016), optionally complemented with snow density samples of the avalanche deposit or by assuming a plausible density to calculate mass (Jamieson et al., 2014). Given current technologies, this is not feasible for all avalanches, and in addition it was not possible at all until a few years ago. Therefore most avalanche inventories rely on size estimates made by humans. Even though avalanches may be identified in remotely sensed imagery with high locational accuracy, there is as of yet no objective way to determine the outlines of avalanches, and – so far – all automatic approaches have been validated against manual mappings (e.g., Lato et al., 2012; Bianchi et al., 2021). Furthermore, suitable remotely sensed imagery is often not available; therefore avalanche outlines are mostly manually mapped, directly in the field or later from photographs.

In practice, field observers or the public may provide an estimation of avalanche size together with the approximate location of the avalanche (a point) or map the outlines of avalanches, while avalanche forecasters recording avalanches may also use photographs provided by third parties for these tasks. It is common practice for avalanches to be assigned a size estimate using a scale. Standardized scales were first proposed about 60 years ago by the U.S. Department of Agriculture (1961) to «provide an effective vehicle for communication between the observers themselves and others» (McClung and Schaerer, 1980, p.15). The earliest classification of avalanches into size categories was based on destructive potential (U.S. Department of Agriculture, 1961). Since then, the classification has been extended and refined by analyzing mass and frequency distributions of avalanches (McClung and Schaerer, 1980). This scale was adopted in several countries (among others Canada and New Zealand). In addition, in the United States the destructive scale is often combined with a relative scale, where avalanches are given a size relative to the avalanche path they occurred on (Birkeland and Greene, 2011). In this system, the size of an avalanche is dependent on its location (McClung and Schaerer, 1980). Both scales use five size classes, with size 1 being the smallest and size 5 being the largest avalanche. With some variations, the destructive scale was adopted in Europe in 2009 and later complemented with more details. An overview of the scales currently used in North America and Europe is shown in Tab. 14.1.

Table 14.1: Definition of avalanche size for Europe (EAWS, 2023) and North America (Canadian Avalanche Association, 2016; American Avalanche Association, 2022, Canadian definition for the description of potential damage, the other parameters are identical). In Europe, length and volume are subsumed under the headline typical dimensions (EAWS, 2023).

Size	Parameter	European definition	North American definition
1	Potential damage	Unlikely to bury a person, except in run out zones with unfavourable terrain features (e.g. terrain traps).	Relatively harmless to people.
	Runout Length	Stops within steep slopes.	—
	Volume	10-30 m	10 m
	Mass/ Impact pressure	100 m <sup>3</sup>	—
2	Potential damage	May bury, injure or kill a person.	Could bury, injure, or kill a person.
	Runout Length	May reach the end of the relevant steep slope.	—
	Volume	50-200 m	100 m
	Mass/ Impact pressure	1'000 m <sup>3</sup>	—
3	Potential damage	May bury and destroy cars, damage trucks, destroy small buildings and break a few trees.	Could bury and destroy a car, damage a truck, destroy a wood-frame house or break a few trees.
	Runout Length	May cross flat terrain (well below 30°) over a distance of less than 50 m.	—
	Volume	several 100 m	1 km
	Mass/ Impact pressure	10'000 m <sup>3</sup>	—
4	Potential damage	May bury and destroy trucks and trains, may destroy fairly large buildings and small areas of forest.	Could destroy a railway car, large truck, several buildings or a forest of approximately 4 hectares.
	Runout Length	Crosses flat terrain (well below 30°) over a distance of more than 50 m. May reach the valley floor.	—
	Volume	1-2 km	2 km
	Mass/ Impact pressure	100'000 m <sup>3</sup>	—
5	Potential damage	May devastate the landscape and has catastrophic destructive potential.	Largest snow avalanches known. Could destroy a village or a forest area of approximately 40 hectares.
	Runout Length	Reaches the valley floor. Largest known avalanche.	—
	Volume	>2 km	3 km
	Mass/ Impact pressure	>100'000 m <sup>3</sup>	—
			100'000 t / 1'000 kPa

Inventories of avalanches mapped either directly in the field or later from photographs have been used in numerous studies (e.g., Hafner et al., 2021; Bühler et al., 2022; Techel et al., 2022) but are known to be incomplete (Schweizer et al., 2020) and biased towards accessible terrain and larger avalanches (Hendrikx et al., 2005). Avalanche size may be directly derived from avalanche outlines (Schweizer et al., 2020; Völk, 2020; Bühler et al., 2019), for example, by thresholding the mapped area. In addition to manual avalanche outline mapping, avalanches have increasingly been mapped (manually or automatically) from remotely sensed imagery such as satellite images or orthophotos acquired from airplanes or drones (e.g., Korzeniowska et al., 2017; Eckerstorfer et al., 2017; Bühler et al., 2019; Bianchi et al., 2021; Hafner et al., 2022). Specifically satellite imagery has the potential to close the information gap in avalanche documentation and record avalanche occurrences over large areas with a consistent methodology, thereby complementing existing databases (e.g., Lato et al., 2012; Vickers et al., 2016; Eckerstorfer et al., 2017; Bühler et al., 2019).

## Data and methods

To explore the reliability of avalanche size estimates and avalanche outline determination, we conducted three user studies, described in Sections 15.1 to 15.3. In all three studies, we simulated different typical size estimation or avalanche mapping tasks based on either oblique photos or remotely sensed images. For each of the three experiments, this translated to the following common task for the study participants: an assessor is given an image and has to (1) detect the avalanche(s) in the image. If an avalanche is detected, the assessor (2) either judges the size of the avalanche or distinguishes between avalanche and no avalanche by drawing an outline.

### 15.1 Study 1: avalanche size estimation

To explore the reliability of avalanche size estimates provided by avalanche practitioners, we developed a survey consisting of 10 photographs of avalanches (see the Supplement C). The photographs used were originally captured for the purpose of documenting avalanches. They show clearly identifiable avalanches in various terrain and were chosen to provide a diverse set of perspectives, sizes and illumination, as well as the presence of various reference objects. In the survey, each participant was asked to estimate the size of each of the 10 avalanches using the five-class integer scale, which we refer to as the «full» size (for instance, size 3; see Tab. 14.1). After estimating the full size of an avalanche, participants had the opportunity to provide an intermediate size («half»-size, nine levels; for instance, size 2.5). As a second task, we asked participants to rate the importance of the factors characterizing avalanche size for their size estimations on a four-point Likert scale as either *very important*, *important*, *less important* or *not at all important* (factors: destructive potential, dimensions, runout and volume; Tab. 14.1). We designed the survey with a European audience in mind and only later decided to extend it to North America. For this reason, runout and volume were included as factors even though they are not part of the North American avalanche size definition. Similarly, in the European definition typical length and volume are presented under the headline typical dimensions (EAWS, 2023), a term which is not present in the North American definition.

The survey was sent to avalanche practitioners, primarily regional avalanche forecasters in Europe and North America, through personal contacts or using forecast center mailing lists. The survey was available in English, French, German and Italian. We asked participants at the beginning of the survey whether they were avalanche forecasters, and in which country they

work. In total, 170 responses were received: 105 from Europe, and 65 from North America. The proportion of professional avalanche forecasters in our dataset was 86% (146). The other 24 participants either had additional roles besides forecasting or worked closely with the avalanche warning service, for example as avalanche educators, mountain guides, ski patrollers or field observers for a warning service. The professional forecasters in our sample were from the United States (39 participants), Italy (33), Canada (17), Norway (15), Spain (10), Austria (10) and Switzerland (7), while all other countries had 2 or fewer participants.

## **15.2 Study 2: avalanche mapping from oblique photographs**

To investigate the reliability in avalanche outlines, we asked nine people, who map avalanches as part of their professional duties, to map six avalanches in the area around Davos (Switzerland) from winter 2020/21.

For each avalanche, we provided three to six photographs and indicated the approximate location by giving the name of a ridge or summit in the proximity of the avalanche (distance 50 to 300 m). Mapping was conducted using operational mapping tools, which provide the user with a topographic map (at best with a scale of 1:10'000; swisstopo, 2020a), orthophotos (resolution of 10×10 cm; swisstopo, 2020b) and slope incline classes for areas steeper than 30° (resolution of 10×10 m). Each participant was asked to map the six avalanches with the same accuracy as they usually would when mapping avalanches. In addition to the nine participants, we used the avalanche outlines that were initially mapped for documentation purposes in winter 2020/21.

To create a reference for the map analysis, we georeferenced one image per avalanche with the monoplottting tool developed by Bozzini et al. (2012, 2013), then drew and exported the avalanche outlines. Since this approach allows for a much more accurate localization of avalanche outlines, we used these as a reference in this study. For one avalanche (a in Fig. 16.4), the deposit was obscured by a tree in the only photograph where the whole avalanche was visible. This part of the avalanche was therefore disregarded in the analyses including this reference.

## **15.3 Study 3: avalanche mapping from remotely sensed imagery**

The third experiment concerned the mapping of avalanche outlines from remotely sensed imagery. In addition to comparing mapped avalanche outlines between individuals, this experiment allowed for exploring some of the potential factors influencing the quality of mapped avalanche outlines (illumination, snow conditions, avalanche type, image resolution; Hafner et al., 2022).

We selected two georeferenced images acquired under different snow conditions (see Tab. 15.1) without artifacts, without saturation and as 16-bit radiometric information. The images were processed in Agisoft Metashape. To obtain different resolutions, we bi-linearly resampled the data in the red-green-blue channel (RGB) to 25 cm, 50 cm, 1 m and 2 m spatial resolution (for native resolutions see Tab. 15.1). For separating illuminated from shaded areas, we used a support vector machine classifier to calculate a shadow mask (see Hafner et al., 2022, for details).

We provided a standardized introduction to the five participants, who were all familiar with avalanches and remotely sensed imagery. All visible avalanches were to be digitized in the software ArcGIS Pro, starting from the coarsest (2 m) and ending with the finest resolution (25 cm). Images with higher resolution were only made available after the mapping of the (one-step) coarser resolution had been completed. Participants could not re-examine their earlier mappings. They had access to the topographic map (at best a scale of 1:10'000; swisstopo, 2020a) and slope incline classes for areas steeper than 30° (resolution of 10×10 m). We instructed participants to outline all areas with signs of avalanche activity rather than drawing

Table 15.1: Properties of remotely sensed imagery, which was used to investigate variations in the performed avalanche mapping.

Acquisition date	Sensor	Mean ground sampling distance (GSD)	Area covered [km <sup>2</sup> ]	Snow and avalanche conditions
16 March 2019	UltraCam Eagle M3 (piloted airplane)	12 cm	2.2	Following a period with numerous dry-snow avalanches
25 February 2021	WingtraOne (drone)	4 cm	0.7	Following a period with numerous wet-snow avalanches

individual events; thus, they were asked to delimit all visible avalanche regions but not to separate them into individual avalanche polygons. The participants did not see the mapped outlines from other participants before they had finished with the highest resolution.

## 15.4 Data analysis

### 15.4.1 Avalanche size estimates

Presumably, having many assessors performing the same task is a rare exception; thus, in most situations only a single estimate for avalanche size is available. Therefore, the reliability of an individual estimate is of interest. Not making an assumption about whether any two size estimates contain the true label, the agreement between raters can be considered an indirect indicator of reliability (Stewart, 2001). For the avalanche size estimation study (Sect. 15.1), we calculated inter-rater agreement as the proportion of agreements in avalanche size between any two raters for the 10 avalanches ( $P_{\text{agree}}$ ). Following Stewart (2001), if random errors between two raters are independent, then the correlation between two raters' estimates cannot be larger than the product of their reliabilities, except by chance. In other words without knowing which rater is more competent or reliable, the reliability ( $rel$ ) of an individual rater is the geometric mean of the individual reliabilities (Techel, 2020, p. 35). In the special case with two raters  $i = \{1, 2\}$ ,  $rel$  can be derived as:

$$rel = \sqrt{rel(1) \times rel(2)} = \sqrt{P_{\text{agree}}(1, 2)} \quad (15.1)$$

Reliability  $rel$  thus provides an indication regarding the certainty related to estimates by individuals (Stewart, 2001, pp. 84-85).

Several studies have shown that the competency of raters influences the reliability of the labels (e.g., Lampert et al., 2016; Wong et al., 2022). We therefore investigated whether some raters more often provided different or rather extreme size estimates compared to others. As we are lacking an independent ground truth label, we infer a ground truth size assuming that the consensus or most frequently chosen size is a suitable approximation. This is a frequently used approach when no ground truth label is available (e.g., Lampert et al., 2016; Wong et al., 2022). Thus, we extracted the mode ( $s_{\text{mode}}$ ) and median size. However, as the number of participants differed between North America and Europe, and not wanting to favor either in case of differences, we considered the mean of the corresponding median size in North America and Europe for avalanche  $j$  as the reference size  $\bar{s}_j$ . In the case that  $\bar{s}_j$  was between two integer values, for instance, 2.5, we considered the result inconclusive and treated the correspondingly lower and higher integer size as correct too (here size 2 and 3). Similarly, in the case of equal



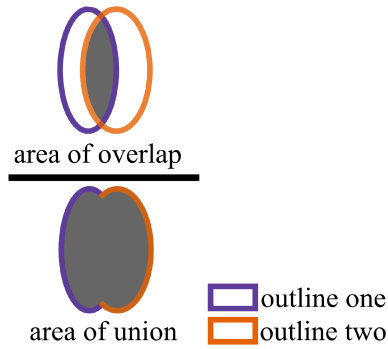


Figure 15.1: Intersection over union (IoU) with the area of overlap (AoO) and the area of union (AoU).

votes for two avalanche sizes, we considered both for the calculation of agreement with  $s_{\text{mode}}$ .

To obtain an indication on the competency of individual raters, we derived a proportion of those that are «correct» for each rater  $i$  of  $P_{\text{correct}}$ , defined as the number of size estimates  $s_{ij}$  being equal to  $\bar{s}_j$  divided by the number of avalanches. As an alternative approach, we calculated  $P_{\text{mode}}$  for each rater  $i$ , specified as the number of size estimates  $s_{ij}$  equal to the most frequently chosen size  $s_{\text{mode}}$  divided by the number of avalanches.

We used the Wilcoxon rank-sum test and the proportion test (as implemented in R Core Team, 2021) to test for significant differences between groups. We considered  $p$  values  $\leq 0.05$  as statistically significant.

### 15.4.2 Avalanche outline determination

For the outline determination exercises (studies 2 and 3), we calculated the intersection over union (IoU) as an indicator of spatial agreement in the mappings by any two annotators (e.g., Levandowsky and Winter, 1971). Here, IoU describes the overlapping proportion of two avalanche areas (AoO) relative to the combined area of the two avalanche areas (AoU):

$$\text{IoU} = \frac{\text{area of overlap (AoO)}}{\text{area of union (AoU)}}, \quad (15.2)$$

where IoU lies between 0 (no overlap) and 1 (full overlap; Fig. 15.1).

We used three variations for IoU:

- $\text{IoU}_{\text{pairwise}}$ , which is the ratio between the intersection of any two individual mappings to the union of these two mappings;
- $\text{IoU}_{\text{all}}$ , which is the ratio between the intersection of an individual mapping to the union of all mappings;
- $\text{IoU}_{\text{ref}}$ , which is the ratio between the intersection of an individual mapping and the reference mapping to the union of these two mappings.

As for the avalanche size estimation study (Sect. 15.1), we explored annotator competence. In study 2 (Sect. 15.2), we used the reference mapping as ground truth. In study 3 (Sect. 15.3), with five participants, we assumed that the area marked as an avalanche by a simple majority (three out of five participants) represented a good approximation of a ground truth.

## Results

### 16.1 Avalanche size estimation (study 1)

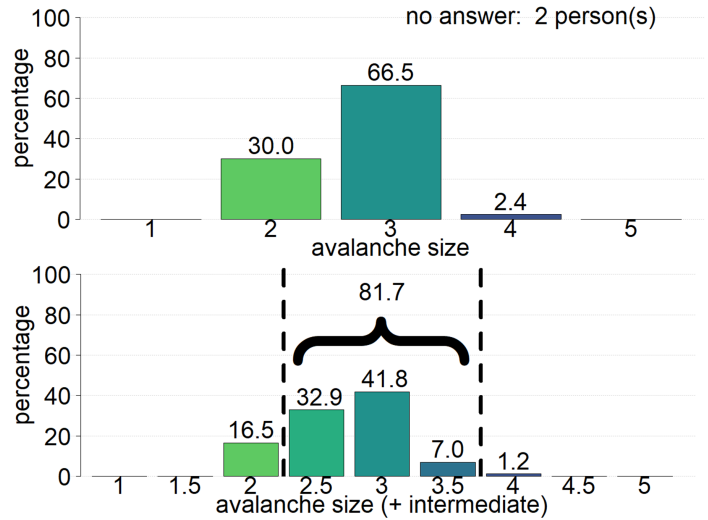
A total of 170 people participated in the survey and estimated the size of 10 avalanches, shown in Fig. 16.1. The **agreement rate between any two size estimates**  $P_{\text{agree}}$  was 0.53, ranging from 0.22 to 0.68 for individual raters. Nine of these raters had an agreement rate lower than the 95th percentile of the 170 participants ( $P_{\text{agree}} \leq 0.39$ ), indicating particularly low correspondence with avalanche size as perceived by others. Each of these nine raters suggested at least for one avalanche a rather «extreme» avalanche size, a size which less than 10% of the participants had chosen. Without these nine raters, the agreement rate was 0.54, which is only marginally higher than the overall agreement. Considering all responses, the mean reliability  $rel$  of individual estimates was 0.72, ranging from 0.47 to 0.82, and, if excluding the nine raters with the lowest agreement with others,  $rel$  was 0.73.

On average, the **agreement with the size considered correct** ( $\bar{s}_j$ ) was  $P_{\text{correct}} = 0.74$ , or, if treating a simple most frequent vote  $s_{\text{mode}}$  as the reference size,  $P_{\text{mode}}$  was 0.66; 16 participants were in full agreement ( $P_{\text{correct}} = 1$ ) with the avalanche sizes considered the most likely size  $\bar{s}_j$ , while the nine raters with the lowest agreement with others also had low values of agreement with  $\bar{s}_j$ . Excluding these resulted in  $P_{\text{correct}} = 0.76$ . In addition to the 66% of the respondents who provided the same size estimate as the most frequently chosen size  $s_{\text{mode}}$ , another 29% chose the second-most popular neighboring size. Thus, in total 92% of all estimates fell into two adjacent size classes highlighting that there was a reasonable consensus on the most likely size(s). Relaxing the definition for agreement even more (as in Moner et al., 2013), 97% of the responses were size  $s_{\text{mode}} \pm 1$ , ranging from 46% for Fig. 16.1b to 98% for avalanche Fig. 16.1f. The average number of different full size classes chosen was 3.7, ranging between two for avalanche Fig. 16.1f and five for avalanche Fig. 16.1j. The latter example means that each of the five size classes were indicated at least once. This shows that even though most votes were in correspondence with one of the two most frequent size classes, at least some estimates regularly deviated strongly from this opinion. An intermediate size class was given in 26% of all cases. The agreement of the intermediate size estimated by a respondent with the most frequently indicated intermediate size  $s_{\text{mode.intermediate}}$  was 0.49, and for  $s_{\text{mode.intermediate}} \pm 0.5$  the agreement was 0.74, while  $P_{\text{correct}}$  was 0.53. The most frequent intermediate size was always between the two most frequent full sizes, underlining that a share of participants differed in their estimates less than a

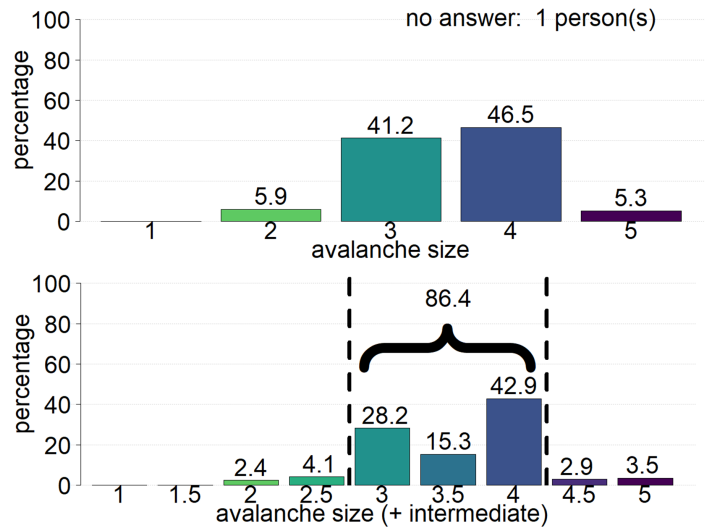
full size (Fig. 16.1). The mean agreement rate  $P_{agree}$ , when allowing for full and intermediate sizes, was 0.37, the reliability  $rel$  was consequently 0.61.



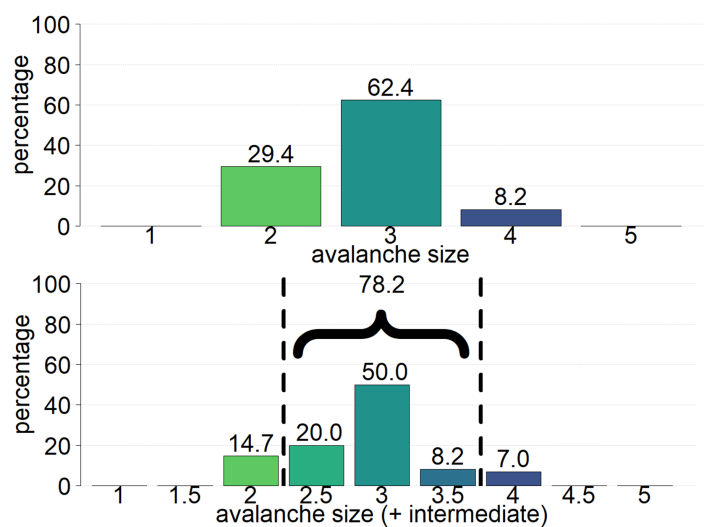
**(a)**  
**Full size:**  
 mode: size 3 (66.5%)  
 median: size 3  
**Intermediate size:**  
 median: size 3

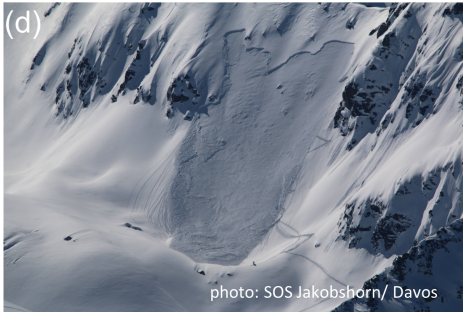


**(b)**  
**Full size:**  
 mode: size 4 (46.5%)  
 median: size 4  
**Intermediate size:**  
 median: size 3.5

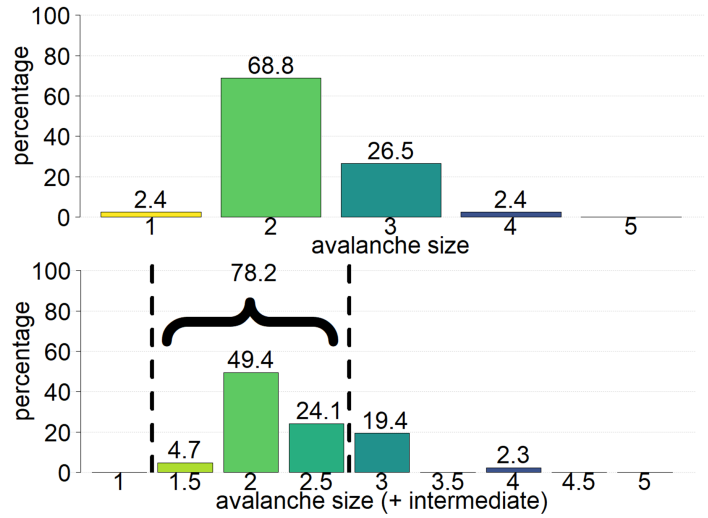


**(c)**  
**Full size:**  
 mode: size 3 (62.4%)  
 median: size 3  
**Intermediate size:**  
 median: size 3

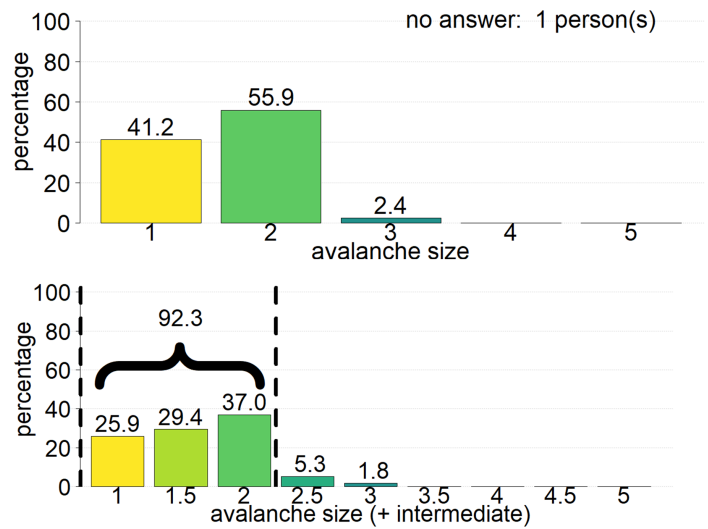




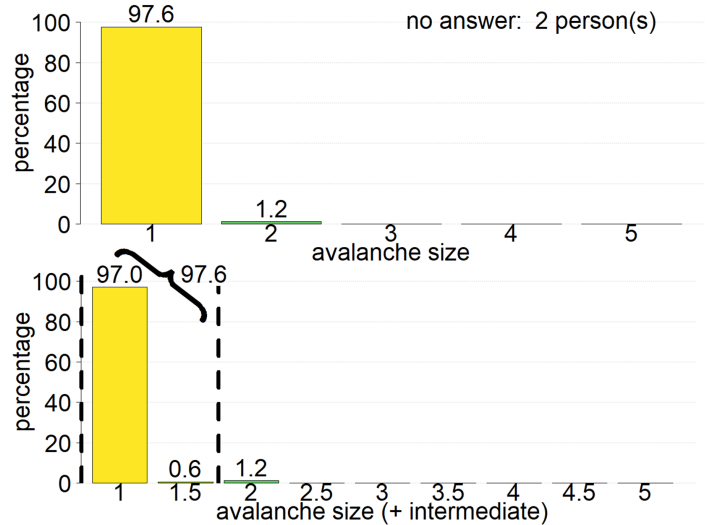
**Full size:**  
 mode: size 2 (68.8%)  
 median: size 2  
**Intermediate size:**  
 median: size 2



**Full size:**  
 mode: size 2 (55.9%)  
 median: size 2  
**Intermediate size:**  
 median: size 1.5

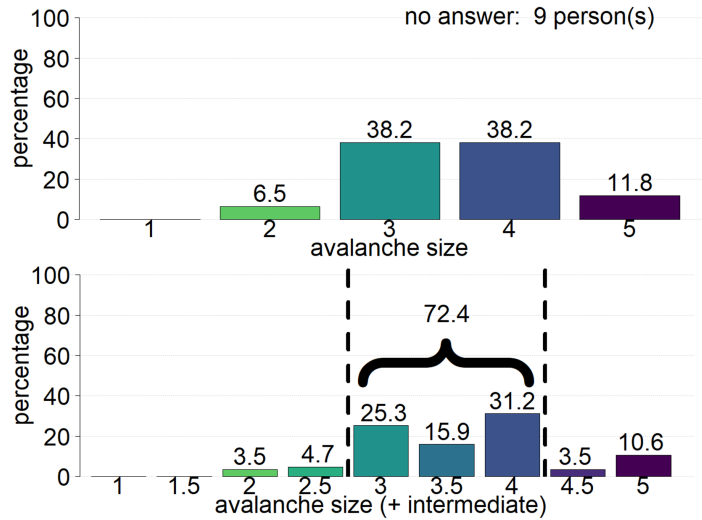


**Full size:**  
 mode: size 1 (97.6%)  
 median: size 1  
**Intermediate size:**  
 median: size 1

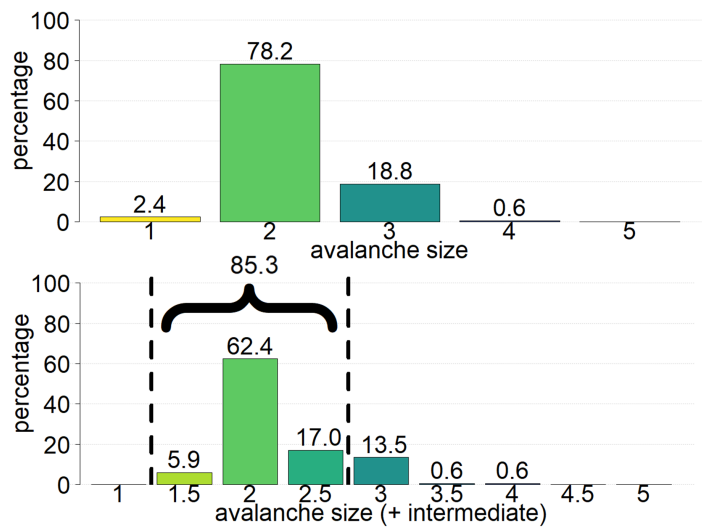




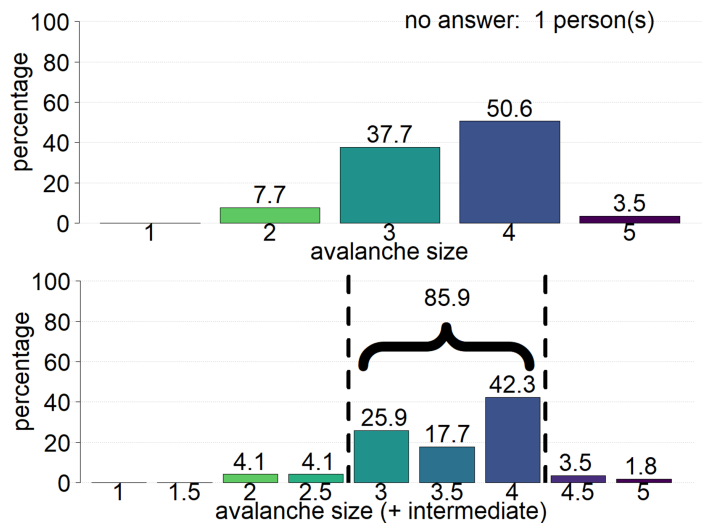
**Full size:**  
 mode: size 3.5 (76.5%)  
 median: size 4  
**Intermediate size:**  
 median: size 3.5



**Full size:**  
 mode: size 2 (78.2%)  
 median: size 2  
**Intermediate size:**  
 median: size 2



**Full size:**  
 mode: size 4 (50.6%)  
 median: size 4  
**Intermediate size:**  
 median: size 3.5





**Full size:**  
 mode: size 3 (59.4%)  
 median: size 3  
**Intermediate size:**  
 median: size 3

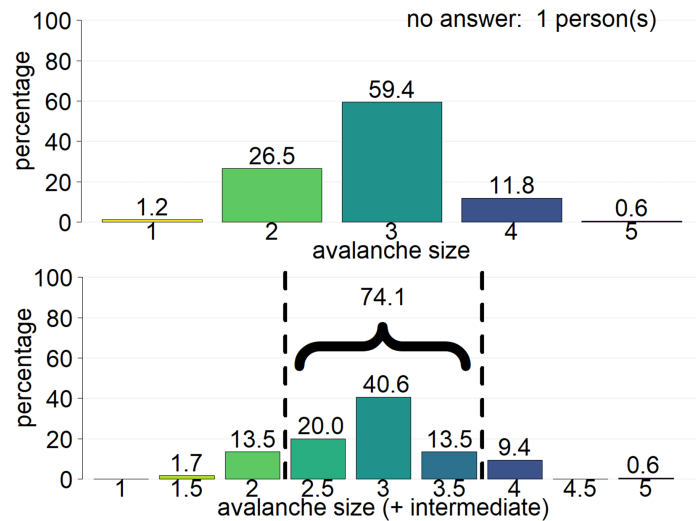


Figure 16.1: Distribution of the size classes and the intermediate sizes assigned to avalanches/pictures (a) to (j) in the survey.

To explore if the size of an avalanche relative within an image and in relation to the surroundings influences size estimation, we included one avalanche twice though the image was cropped and flipped (avalanches Fig. 16.1a and c); 168 out of 170 participants rated both avalanches. Of those 168, 78% indicated the same size, and 15% rated the avalanche one size larger in the close-up view in Fig. 16.1c than in the overview in Fig. 16.1a, whereas 7% rated the avalanche one size smaller in the close-up view compared to the overview. The shift in the proportions is statistically significant (proportion test  $p = 0.036$ ).

When **comparing the results from Europe and North America**, we found the agreement of individual raters in the most frequently estimated size  $s_{mode}$  to be identical ( $P_{mode} = 0.66$ ). This approach slightly favors European respondents, as these contributed a larger share of responses (Europe:  $N = 105$ , North America:  $N = 65$ ). Considering  $\bar{s}_j$  instead, the agreement  $P_{correct}$  was 0.74 overall and 0.66 for both Europe and North America individually. The higher overall agreement results from the definition of the reference size  $\bar{s}_j$ , where two sizes were considered correct if the reference size was located in between two values. North Americans had a tendency to assign smaller sizes than their European counterparts. This is most notable for the three largest avalanches (avalanches in Fig. 16.2b, g and i), with a median size of 4 by Europeans and a median size 3 by North Americans. With the exception of avalanches e and f, differences in size estimates were statistically significant (Wilcoxon rank-sum test:  $p$  ranging from 0.045 to  $< 0.001$ ). Within their continents, respondents had a similar agreement with each other (proportion test:  $p = 1$ ): on average,  $P_{agree}$  was 0.53 within Europe and 0.56 within North America, resulting in  $rel$  of 0.73 and 0.75, respectively. Intermediate sizes, which are more commonly used in North America, were chosen in 31% of the cases by North Americans compared to 23% by Europeans (proportion test:  $p = 0.368$ ). When using intermediate sizes, the agreement with the mode intermediate size,  $s_{mode.intermediate}$  was 53% for Europe ( $\pm 0.5$ : 75%) and 49% for North America ( $\pm 0.5$ : 81%).

Among the **factors used to determine avalanche size** (Tab. 14.1), runout was considered the most important with 56% of respondents considering this factor as very important, followed by volume (very important: 39%), dimensions (29%) and destructive potential (20%)(Fig. 16.3).

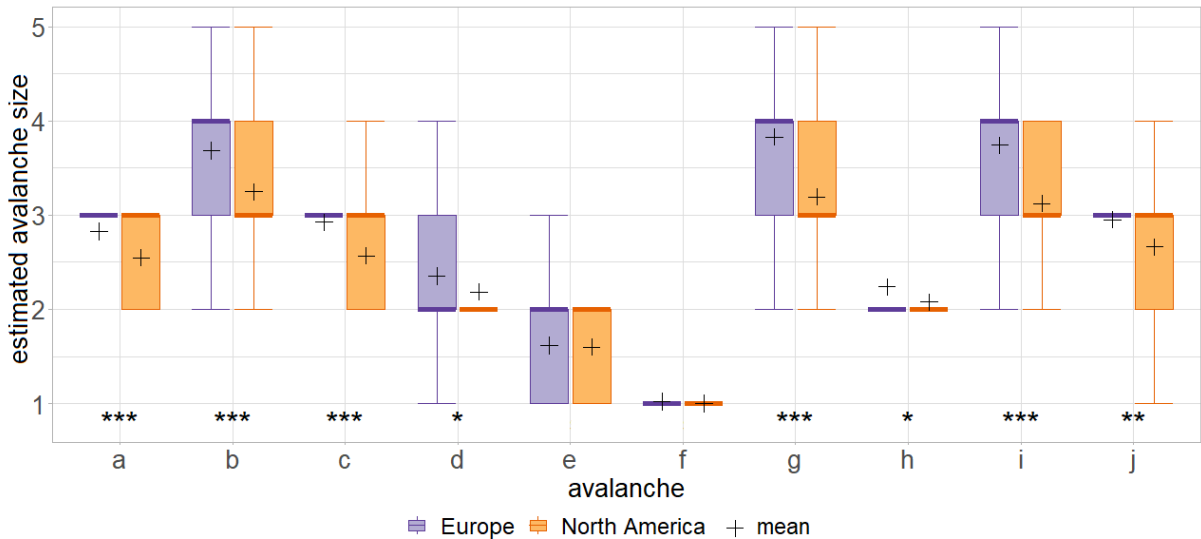


Figure 16.2: Boxplots showing the size distributions for the 10 avalanches for Europe and North America. Mean values are indicated with +. Avalanches are labeled according to Fig. 16.1, and (a) and (c) depict the same avalanche. The results from the Wilcoxon rank-sum test indicate that the differences in avalanche size estimation between Europe and North America are significant for 8 out of the 10 avalanches (all except (e) and (f); \* (0.01, 0.05], \*\* (0.001, 0.01], \*\*\*  $\leq 0.001$ )).

Comparing responses from Europe and North America, we found the most frequent response to be identical for all four factors (very important for runout, important for the other three). However, runout was considered significantly less often as very important in Europe (46%) compared to North America (72%, proportion test:  $p = 0.001$ ). The factor volume showed a similar pattern with significantly more votes from North America (46%) than Europe (18%) for being very important (proportion test:  $p < 0.001$ ) and the opposite pattern for volume being either less important or not at all important (Europe: 39%, North America: 7%, proportion test:  $p < 0.001$ ). The differences between the continents for rating the importance of destructive potential and dimensions were not significant (proportion test with rating of very important:  $p = 0.260$  and  $p = 0.718$ ).

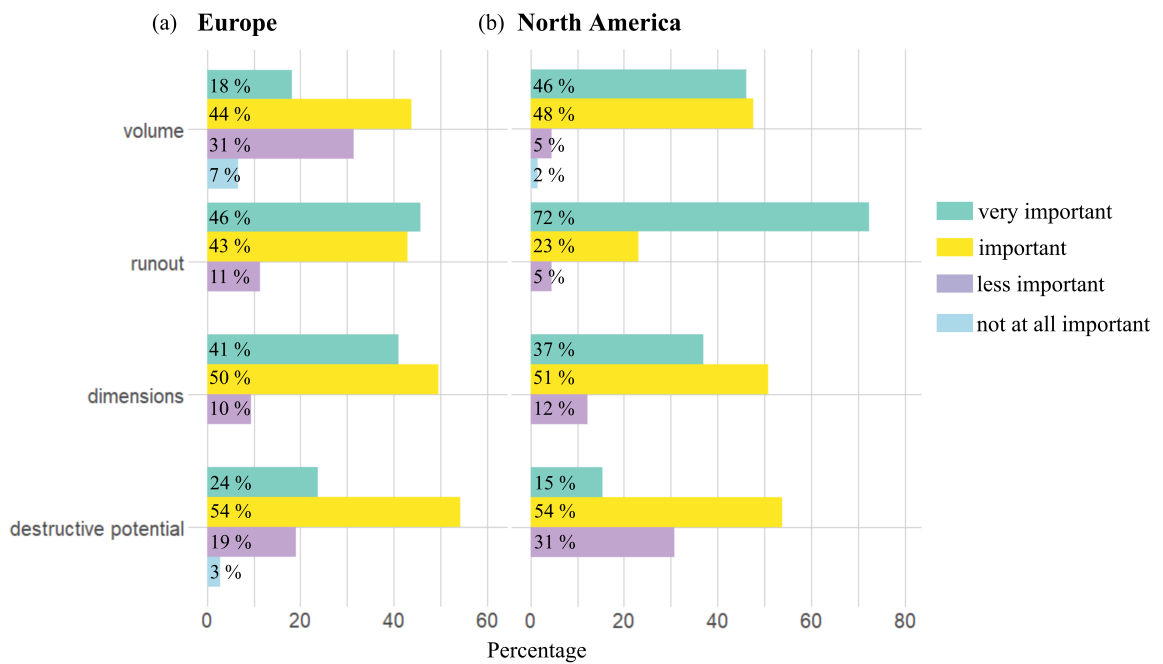


Figure 16.3: Comparison of the importance ranking for the factors determining avalanche size for (a) Europe and (b) North America.



## 16.2 Avalanche mapping from oblique photographs (study 2)

The 10 participants, all very familiar with the study area, centered the avalanches based on oblique photographs around the corresponding reference mapping (Fig. 16.4), identifying the correct locations.

On average, the overlapping proportion of the mappings of any two participants of  $IoU_{\text{pairwise}}$  was 0.52, varying from 0.32 for the worst pairwise agreement to 0.69 for the best one (Tab. 16.1). Individual pairwise comparisons are shown in Fig. B.1 in the Appendix. When comparing individual mappings to the area mapped by at least one person as AoU, the mean  $IoU_{\text{all}}$  is 0.31, ranging from 0.21 to 0.41. Only a fraction of 9% of the combined area of union (AoU for all 10 participants) was identified by all participants as an avalanche (area of overlap for all 10 participants, AoO) showing the considerable scatter of individual mappings. Comparing individual mappings to the reference resulted in a mean  $IoU_{\text{ref}}$  of 0.60, with a minimum of 0.40 and a maximum of 0.80. These areas of higher agreement between participants, visible in darker hues in Fig. 16.4, coincide with the outlines from the reference mapping highlighting that variations happened around the reference. In other words, individual mappings had a higher correspondence with the reference mapping compared to mappings by other individuals (see also Fig. B.1 in the Appendix). The large variation between individual mappings was also shown when analyzing the absolute values of the mapped areas (Tab. 16.2): the largest mapped area was between 2 and 4 times larger than the smallest mapped area (Fig. 16.4f and b). Additionally, the comparison with the reference showed a systematic tendency towards underestimation of the area, as in all cases the median mapped area was between 10% and 36% smaller than the reference area.

Two of the raters had statistically lower pairwise overlap in their area of avalanche activity (mean  $IoU_{\text{pairwise}} \leq 0.49$ ) compared to the other eight raters (mean  $IoU_{\text{pairwise}} \geq 0.69$ , Wilcoxon rank-sum test:  $p = 0.011$  and  $p = 0.002$ ; see Fig. B.1 in the Appendix). These two raters also had the lowest agreement with the reference mapping ( $IoU_{\text{ref}} \leq 0.44$ ), lower than the other eight ( $IoU_{\text{ref}} \geq 0.53$ ).

Table 16.1: Intersection over union (IoU) for avalanches mapped from oblique photographs (study 2). Values represent the mean of six avalanches.

	$IoU_{\text{pairwise}}$	$IoU_{\text{all}}$	$IoU_{\text{ref}}^*$
Mean	0.52	0.31	0.60
Min	0.32	0.21	0.40
Max	0.69	0.41	0.80

\*without deposit from Fig. 16.4a

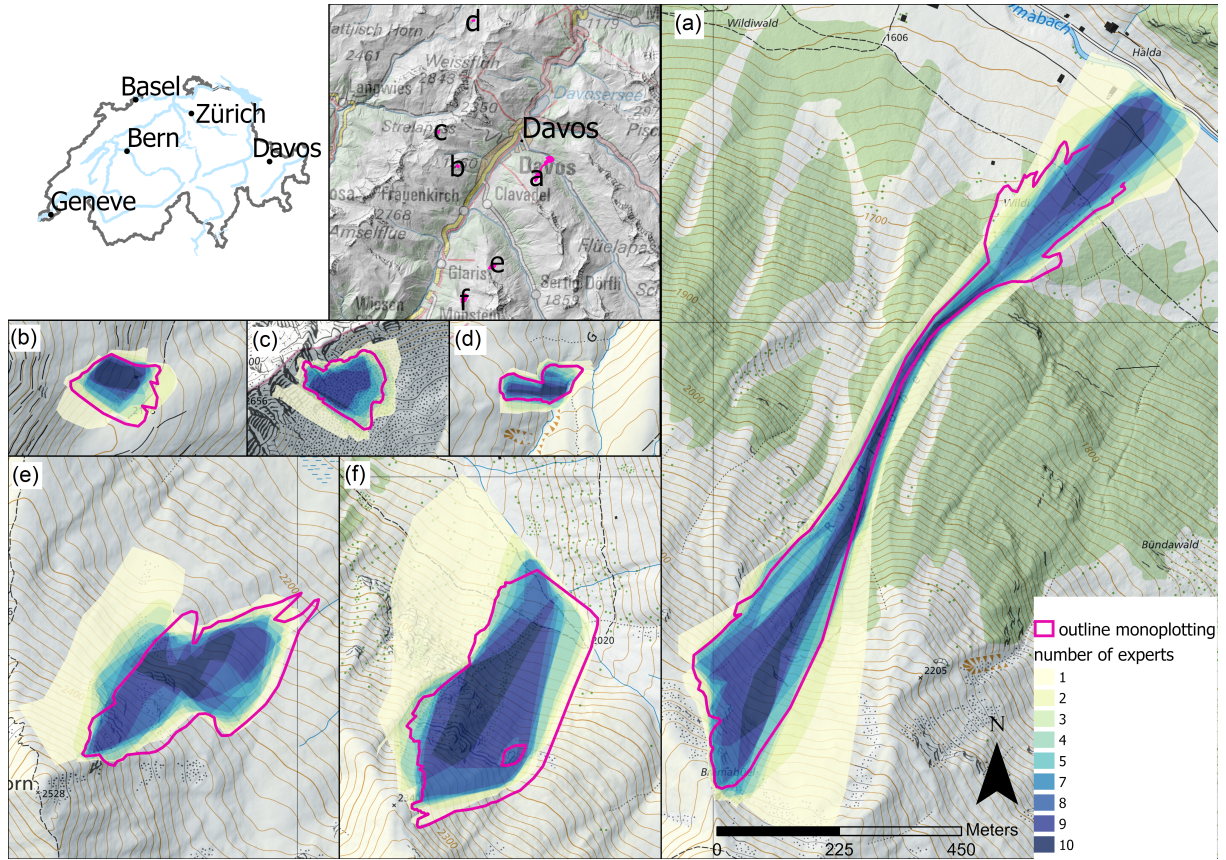


Figure 16.4: Heat map illustrating expert agreement on avalanche area for the six avalanches mapped from oblique photographs. Dark blue indicates areas of very good agreement, identified as part of an avalanche by all 10 experts. For location and size comparison the outlines of the avalanches, mapped from the photographs georeferenced with the monoplotting tool by Bozzini et al. (2012, 2013), are shown as a reference in pink (for avalanche a the lower part was occluded by a tree; map source: Federal Office of Topography).

Table 16.2: Avalanches mapped from oblique photographs (numbering corresponds to Fig. 16.4, study 2). Shown are the areas of the reference mapping, and the respective median, minimum and maximum of the 10 individual mappings. The relative difference to the reference (in %) is indicated in parentheses.

Avalanche	Reference ( m <sup>2</sup> )	Median ( m <sup>2</sup> )	Min ( m <sup>2</sup> )	Max ( m <sup>2</sup> )
(a)*	118'615	106'566 (-10)	66'143 (-44)	190'823 (+61)
(b)	13'673	8'741 (-36)	3'745 (-73)	15'518 (+14)
(c)	13'082	11'071 (-15)	7'422 (-43)	16'221 (+24)
(d)	6'570	5'422 (-18)	2'183 (-67)	7'533 (+15)
(e)	63'807	50'127 (-21)	24'804 (-61)	64'680 (+1)
(f)	90'400	71'967 (-20)	58'383 (-35)	114'404 (+23)

\* without lower part of deposit

### 16.3 Avalanche mapping from remotely sensed imagery (study 3)

When visually comparing the mappings, differences can be observed not only between image resolutions (Fig. 16.5) but also between participants (Fig. 16.6). The mean of the pairwise overlapping proportion of avalanches of  $IoU_{\text{pairwise}}$  increased with increasing image resolution from 0.46 at 2 m resolution to 0.68 at 25 cm resolution (Tab. 16.3). Considering the area classified as an avalanche by three or more raters as the reference showed an increase in  $IoU_{\text{ref}}$  from 0.64 at 2 m resolution to 0.80 at 25 cm resolution (Tab. 16.4). Regarding the influence of illumination conditions, all  $IoU_{\text{pairwise}}$  scores were higher in illuminated areas compared to shaded areas of the image (for instance, at 25 cm resolution - illuminated: 0.77, shaded: 0.54; Tab. 16.3). Snow conditions also influenced the agreement of the mappings (Figs. 16.5 and 16.6): the mean  $IoU_{\text{pairwise}}$  was higher in wet-snow conditions (25 cm resolution: 0.90) compared to dry-snow conditions (25 cm resolution: 0.66; Tab. 16.3). Moreover, individual mappings were also much more similar in wet-snow conditions compared to dry-snow conditions with the variations in  $IoU_{\text{pairwise}}$  ranging for dry-snow conditions at 25 cm resolution between 0.56 and 0.77 (mean: 0.66; standard deviation: 0.07), and for wet-snow conditions between 0.88 and 0.91 (mean: 0.9; standard deviation: 0.01; Fig. 16.7d). Overall, variations in mean  $IoU_{\text{pairwise}}$  were smaller across resolutions (0.02 to 0.22) than the differences between the minimum and maximum  $IoU_{\text{pairwise}}$  within one resolution (0.20 for 25 cm resolution to 0.43 for 2 m resolution). This is especially pronounced for dry-snow conditions (Fig. 16.7). The large variations between different experts are also reflected in the avalanche area that was consistently identified by one person over all four spatial resolutions (dark red in Fig. 16.6).

One of the five participants had a lower pairwise agreement compared to the other four (for instance at 2 m resolution:  $IoU_{\text{pairwise}} \leq 0.41$  vs.  $IoU_{\text{pairwise}} 0.37 - 0.76$ ), although this was not significant (Wilcoxon rank-sum test: 2 m resolution  $p = 0.088$ , 25 cm resolution  $p = 0.055$ ). Considering the area classified as avalanche by three or more raters as the best approximation of a ground truth, the mean agreement with this mapping ranged between  $IoU_{\text{ref}} = 0.64$  and  $IoU_{\text{ref}} = 0.80$  (Tab. 16.4). Again the same participant had the lowest mean agreement.

Table 16.3: Mean  $IoU_{\text{pairwise}}$  for different subsets and spatial resolutions.

Subset	Image resolution				Area ( km <sup>2</sup> )
	2 m	1 m	50 cm	25 cm	
Overall	0.46	0.61	0.66	0.68	2.9
Illuminated	0.51	0.67	0.73	0.77	1.6
Shaded	0.36	0.49	0.54	0.54	1.3
Dry snow	0.44	0.59	0.64	0.66	2.2
Wet snow	0.70	0.81	0.86	0.90	0.7

Table 16.4: Mean  $IoU_{\text{ref}}$  for all spatial resolutions.

$IoU_{\text{ref}}$	Image resolution			
	2 m	1 m	50 cm	25 cm
	0.64	0.76	0.79	0.80

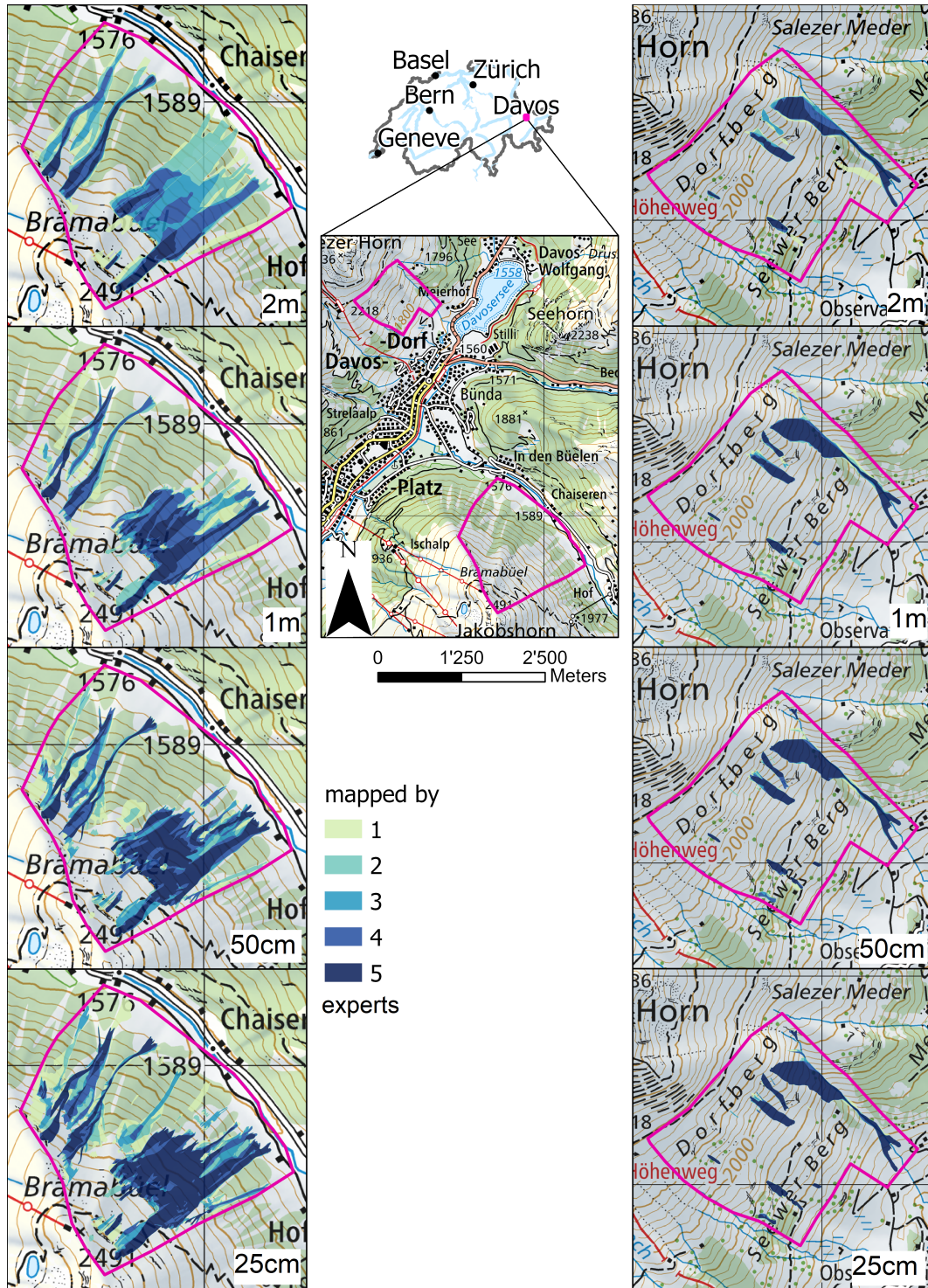


Figure 16.5: Heat map illustrating expert agreement on the avalanche area mapped from remotely sensed imagery for four spatial resolutions (2 m to 25 cm, rows, from top to bottom) for the example dry-snow conditions (left column) and wet-snow conditions (right column). The darker the hue, the greater the agreement of the five experts on the existence of an avalanche in that particular location (map source: Federal Office of Topography).

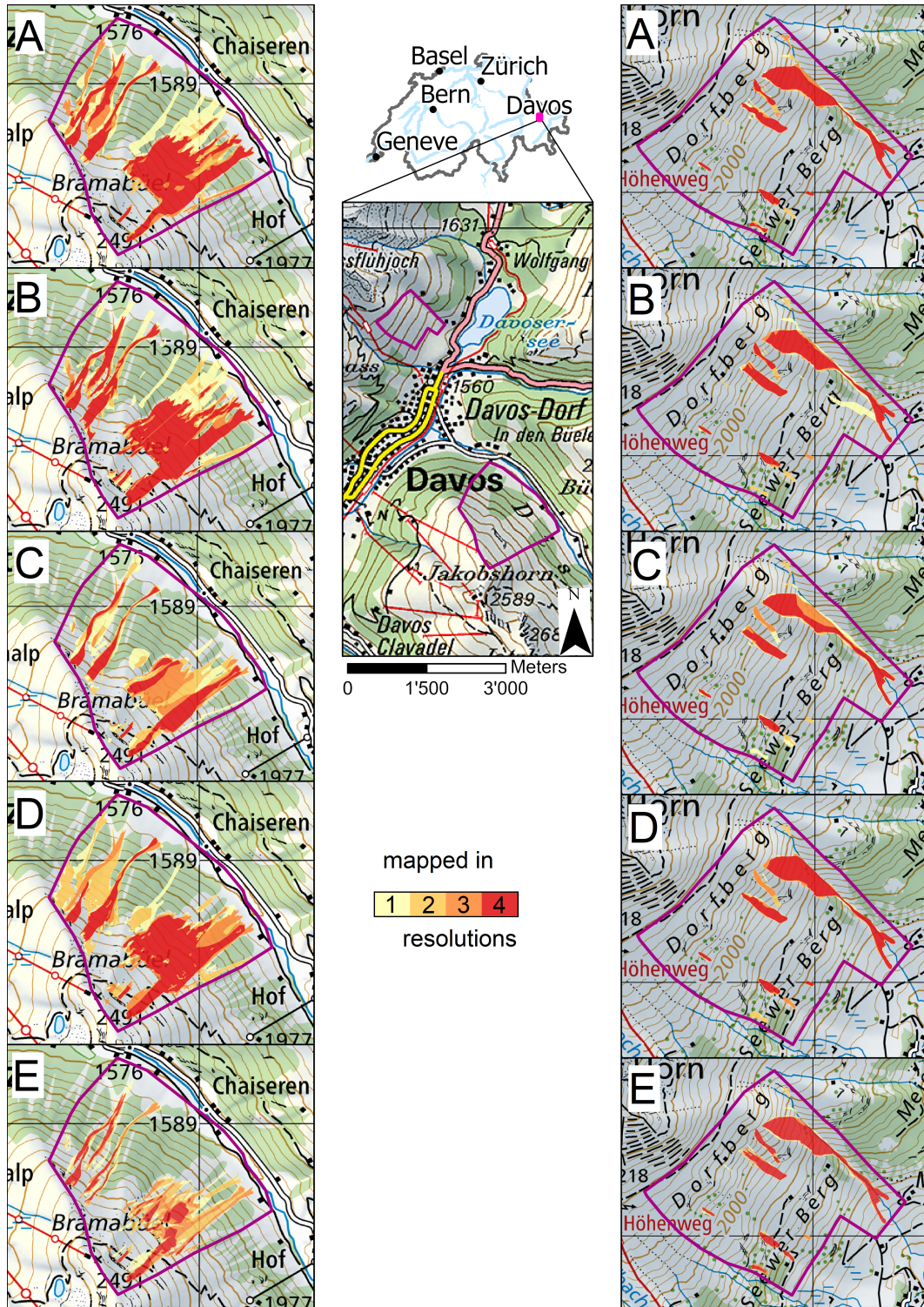


Figure 16.6: Heat map showing differences in the avalanche mappings for participants A to E (rows, from top to bottom), as a function of the four resolutions for the example dry-snow conditions (left column) and wet-snow conditions (right column). Dark hues (red) indicate areas, where an avalanche was detected in all four resolutions; light hues indicate areas where an avalanche was detected in only one resolution (map source: Federal Office of Topography).

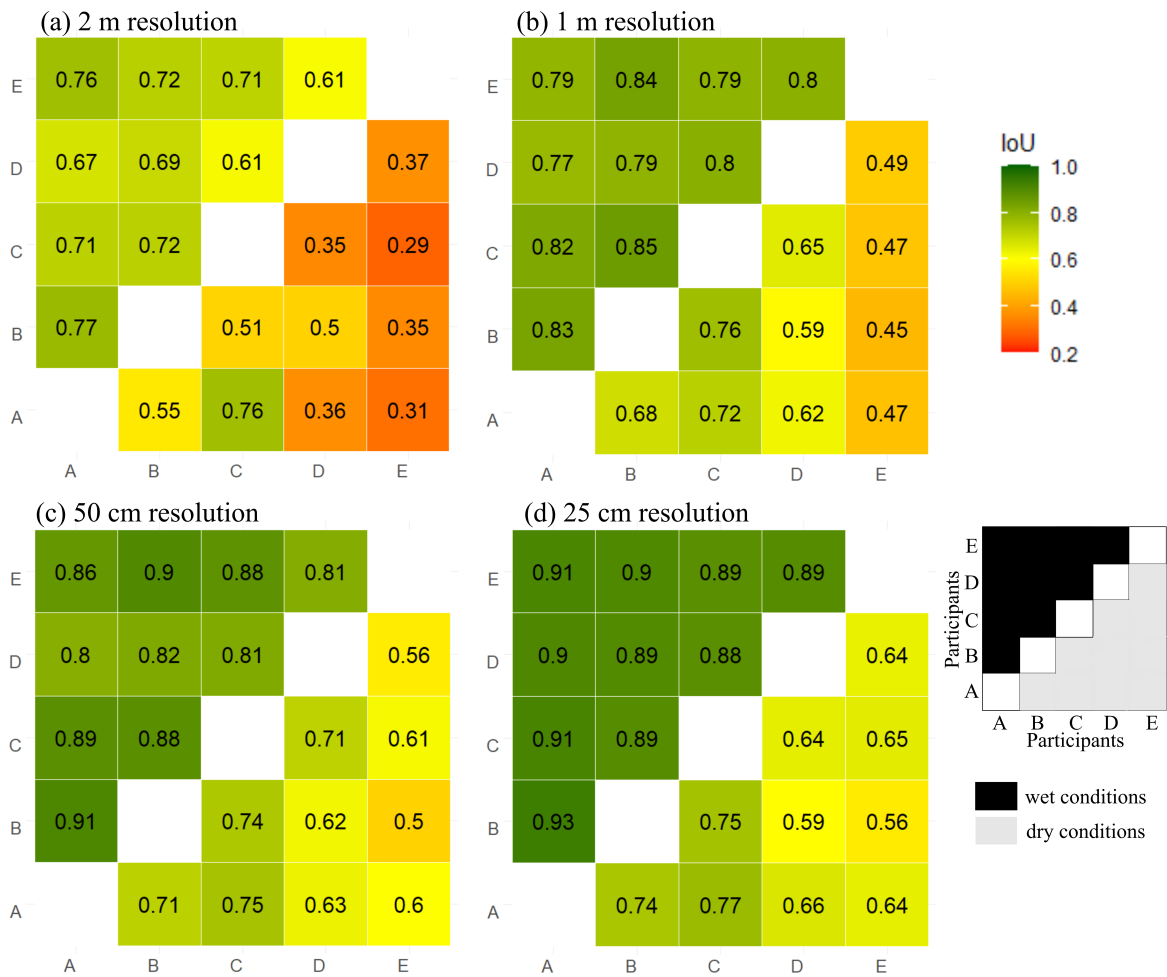


Figure 16.7:  $IoU_{\text{pairwise}}$  for dry- (below diagonal) and wet-snow conditions (above diagonal). The letters A to E represent the different participants, the four tiles (a-d) represent the four resolutions.

## Discussion

We explored the reliability of estimates of avalanche size and detecting the outline of avalanches from images. The key findings are as follows.

- The agreement rate  $P_{\text{agree}}$  between any two size estimates was 0.53, resulting in a reliability  $rel$  of 0.72, while the agreement with the avalanche size considered correct was 0.74, with the most frequently chosen size of 0.66 (mode).
- Significant differences were observed between Europe and North America, both for rating avalanche size and for weighing the factors determining avalanche size.
- The mean overlapping proportion of any two avalanche mappings of  $IoU_{\text{pairwise}}$  was 0.52 (Study 2) and between 0.46 and 0.68 (study 3) and thus lower than the mean agreement with the reference of  $IoU_{\text{ref}}$ , which was 0.60 (study 2) and between 0.64 and 0.8 (study 3).

In the following, we discuss these results by considering definitions, the conclusiveness of the data for the task at hand and the competence of participants. Finally, we provide recommendations for practice.

### 17.1 Avalanche size estimation (study 1)

Our results show that it is difficult to achieve consistent size estimates of avalanches: in only 53% of the cases did any two size estimates agree, and in 66% of the cases an individual estimate agreed with the most frequent size among the respondents (mode). It showed, however, that in most cases disagreements were comparably small with 92% of individual estimates being either equal to the mode  $s_{\text{mode}}$  or equal to the second-most frequent neighboring size or, if considering intermediate sizes, that 74% of the estimates were within one intermediate size class. Comparing our results to previous studies investigating agreement for avalanche size estimates (Tab. 17.1), the agreement rate with the most frequent size ranged from 62% (Moner et al., 2013) to 84% (Hafner et al., 2021). The high agreement rates in Jamieson et al. (2014) and Hafner et al. (2021) are probably related to the fact that these studies relied on a small number of experienced practitioners with a comparably similar background and training. In contrast, both the studies by Moner et al. (2013) and our study included participants from numerous countries and thus respondents with different avalanche backgrounds, leading to a more diverse group of avalanche practitioners. Moreover, Moner et al. (2013) speculated that the changes introduced in the avalanche size definitions shortly before their survey may have lowered the agreement. The

reliability of an individual size estimate in this study was 0.73, highlighting the uncertainty associated with these data. Thus, using size estimates by an individual as ground truth when developing or evaluating models, a perfect model can achieve more than 73% accuracy only by chance if the errors a model makes are independent from the errors contained in the avalanche size labels.

In our survey, we found the lowest agreement with  $s_{\text{mode}}$  for the three largest avalanches ( $\bar{s}_j \geq 3.5$ ) (avalanches Fig. 16.1b, g and i). For these three avalanches,  $s_{\text{mode}}$  differed between Europe and North America. Jamieson et al. (2014) argued that practitioners have more experience with smaller avalanches (sizes 1, 2 and 3), which are much more frequent than larger avalanches, which may cause size estimates of large avalanches to be more variable and less accurate.

Table 17.1: Comparing the agreement in the mode for avalanche size estimates with previous studies.

Study	Average agreement [%]			Raters (samples)
	Full size	Full size $\pm 1$ class	Intermediate size	
Moner et al. (2013)*	62	—	25	61 (18)
Jamieson et al. (2014)	79	100	44	22 (18)
Hafner et al. (2021)	84	—	—	2 (351)
This study	67	97	49	170 (10)

\* European forecasters only, for the Canadian ones see Jamieson et al. (2014)

We noted systematic differences between size estimates provided by North Americans and by Europeans, with North Americans tending towards smaller sizes (Fig. 16.2). This might stem from differences in the European and the North American definitions (see Fig. 16.2): for the typical length the European definition provides a range, whereas in the North American definition a typical value is given. The European definition encompasses the North American values describing typical length for the smaller avalanches, while it coincides with the upper bound for size 4 and provides only a minimum value for size 5. Another difference is that the North American definition includes typical mass, in line with the definition introduced by McClung and Schaerer (1980), while in Europe typical volume is defined. Combining deposit volume with density measurements or density estimates of the deposit, mass may be determined (mass = volume  $\times$  density). For instance, calculating the mass of avalanches assuming a mean density of  $390 \text{ kg m}^{-3}$ , measured from 95 avalanches at Rogers Pass, British Columbia (McClung and Schaerer, 1985), avalanches are almost 4 times larger in the European compared to the North American definition: 4 times larger corresponds approximately to a half size (e.g., for size 2 the mass according to the definition is  $10^2 = 100 \text{ t}$ ; for size 2.5, it is  $10^{2.5} = 316 \text{ t}$ ). Consequently, the significant intercontinental differences may be, at least partially, attributed to the differing size class definitions.

We also observed differences in the importance ranking assigned to the factors determining avalanche size, with both the criteria runout and volume being considered more relevant for size estimation by North Americans compared to Europeans. This is particularly noteworthy as neither a description of runout nor an indication of volume are part of the North American size definitions (Tab. 14.1). We found the size-determining factor of destructive potential to be considered the least important by North Americans and the second-least important by Europeans. This is surprising as both the definitions in Europe and North America state that avalanche size is classified according to destructive potential (e.g., EAWS, 2023; CAA, 2023).



Furthermore, this finding is also contradictory to that in the study by Moner et al. (2013), where destructive potential was the highest-rated factor. One possible reason for the low importance of destructive potential in our study might be related to the study design, where supplementary information about damage to property or people, beyond the photographs, was absent. Thus, destructive potential had to be inferred from avalanche properties like width, length and volume, which is arguably the normal situation when estimating avalanche size.

To find out if the way an avalanche is shown in an image influences size estimations, we included one avalanche twice, changing the perspective and zoom. Even though 78% rated the avalanche in Fig. 16.1a and c the same, we observed a significant proportion of larger estimates in the close-up view. We suspect that this might be caused by the perception of the avalanche being larger when covering more area in the photograph. But our sample is small, and understanding the effect of perspective and area covered by avalanche would require further investigations with a more meaningful sample.

## **17.2 Avalanche mapping from oblique photographs (study 2) and from remotely sensed imagery (study 3)**

Study 2 required participants to first find the location of the avalanche on the map, matching the topography visible in the images with the topography as shown on the map, before mapping was possible. The 10 experts, all very familiar with the study area, located the avalanches in the same place. This first step of the assignment would have been more difficult for someone that did not know the area well, possibly resulting in entirely different locations and hence mappings. Thus, the mean overlapping proportion of any two mappings ( $IoU_{\text{pairwise}}$ ), which was 0.52 in our study, may potentially be 0 if an avalanche is located in the wrong place. We therefore assume that an  $IoU_{\text{pairwise}}$  of 0.52 may well describe the upper limit of agreement in mappings from oblique photographs.

Study 2 allowed for a comparison with a reference mapping using a methodology superior to the approach the 10 experts used. Thus, the agreement between experts' mapping and the reference mapping can be interpreted as the proportion correct and, hence, allows for assessing the experts' competence in mapping avalanche outlines. The overall proportion considered correct ( $IoU_{\text{ref}}$ ) was 0.6, with a clear bias towards smaller mapped areas compared to the reference. The results showed that experts were not equally competent, with the proportion correct ranging from 0.4 to 0.8 (Tab. 16.2). The agreement between individual mappings and the reference mapping is larger than the agreement in the mappings between participants ( $IoU_{\text{pairwise}} = 0.52$ ). This means that the reliability of individual mappings would be underestimated when relying on a measure like the agreement rate between domain experts. If competence is known, it would be possible to weigh individual mappings if two or more mappings were available, likely resulting in more reliable results. Overall, we consider the mapping of avalanches using oblique photographs a challenging task to perform consistently and accurately.

In study 3, five participants had to identify the avalanches in the remotely sensed imagery, for each of four image resolutions, before mapping them. In other words, whether a point is identified as an avalanche is a combination of existential and extensional uncertainty (Molenaar, 1998), addressing the following questions: is there an avalanche, and where are the boundaries? This uncertainty was lower with higher image resolution and for illuminated compared to shaded parts of the image, it allowed participants to identify avalanche area more consistently, and confirmed the findings of earlier work (Hafner et al., 2022). Moreover, snow conditions influenced agreement too, with higher agreement under wet-snow compared to dry-snow conditions (Fig. 16.7). We suspect that this difference is caused by the presence of liquid water in the case

of a wet snowpack, which leads to more pronounced avalanche boundaries compared to dry-snow conditions.

Another important finding from these studies are the large differences in the areas mapped as an avalanche by the experts. For the six avalanches in study 2, the largest mapped area was between 2 and 4 times larger than the smallest mapped area (Tab. 16.4). For study 3, variations across resolutions were found to be smaller than the variations in  $IoU_{\text{pairwise}}$  per resolution, suggesting individual experience and competence have a larger impact than the underlying spatial resolution. If avalanches would be classified automatically using area or by extracting width and length from the mapping (e.g., Schweizer et al., 2020), completely different size classes may result due to these variations. For instance, comparing the mappings of 4'000 avalanches with the reported size estimate, Völk (2020) showed that the median area of size-2 avalanches was about 3 to 5 times larger than size-1 avalanches, or that the mapped area of size-4 avalanches was about 7 times larger compared to size 3 avalanches (Völk, 2020, pp. 49, 51). Comparing these values to the variation observed in the mappings by different experts in our study suggests that estimating avalanche size based on mapped area would, quite frequently, result in different size estimates.

### 17.3 Implications for practice

Several sources of error may impact the reliability of tasks involving human judgment and estimation. These are related to the data being suitable; the skill or competence of the rater interpreting the data; and, finally, the level of generalization (e.g. forcing avalanches into five size classes). In the context of size estimation and outline determination, low image resolution, lack of image context/reference objects, and unfavourable illumination conditions such as shade or diffuse light may impact the conclusiveness of the data for the task to be performed and, hence, reliability. Similarly, differences in the raters' backgrounds and experiences may lead to different levels of competence in interpreting the data. Furthermore, variations in the perception and definition may result in systematic variations in size estimates or outlines. Moreover, the binary choice between an avalanche and no avalanche, when mapping avalanche outlines, does not allow for expressing uncertainty. This may lead to more pronounced differences in the avalanche area identified by participants. Addressing and being aware of these issues will improve the quality of avalanche size estimation and outline determination.

The results of this study indicate that size inventories from North America and Europe, different warning services within Europe, or simply different domain experts may systematically differ in their assessments. Consequently, transferability of size inventories between continents or different warning services may be limited. To achieve a common understanding and comparable size estimation, in particular for expert forecasters and for observers, we suggest a joint effort of the continental and national avalanche associations, together with avalanche forecasters and other avalanche practitioners, to develop training tools that help standardize the size estimations. An in-depth analysis of current protocols or training programs could be fruitful and serve as a first step in tackling this issue. One option might be training people in the «all observables approach» advocated by McClung and Schaerer (1980), which requires imagining the objects that might be destroyed in the track or start of the runout zone of an avalanche.

In the meantime, the uncertainty related to size estimates may be reduced taking into account second estimates and/or jointly deciding on the size in the case of disagreements (e.g., Hafner et al., 2021). Additionally, we recommend the use of intermediate sizes in the following way: first the (full) size class should be estimated. In a second step, the assessor may judge whether avalanche size is low or high or in the middle of the class (like suggested for evaluative social

judgments by Goffin and Olson, 2011). If it was low or high, this would result in the intermediate sizes between the chosen and the upper or lower adjacent full size. Practically, this could mean that from a full size 2 the assessor may, in a second step, assign size 1.5 if the avalanche is at the lower end of size 2, assign 2.5 if it is at the upper end or keep a full size 2 if it is a «typical» avalanche for that size.

While the observed variations in avalanche outlines may partly be attributed to a different background and level of experience, we argue that it is partly caused by the lack of a common, precise definition of where to delimit an avalanche. We are not aware of any unambiguous, actionable guideline of where exactly to place the visible outline. Arguably, there is no “natural”, self-evident definition, especially for dry-snow avalanches. Consistency, in the sense of repeatability across expert annotators, can perhaps only be achieved through a generally agreed consensus that includes shared but, to some degree arbitrary, conventions. It appears that a standardization effort may be beneficial and that standardized training could go a long way towards reducing the spread between different experts and organizations; even if some causes of variability, e.g. lighting conditions after a large snowfall, cannot easily be controlled and will remain.

If reliable mappings from photographs are required we recommend second mappings, jointly deciding on the extent of the outline, using a monoplotted tool (e.g., Bozzini et al., 2013) or the overlay image capabilities of Google Earth. For remotely sensed imagery we advocate using a spatial resolution of 50 cm or finer for the detailed segmentation of specific avalanches, whereas a resolution of approximately 2 m seems to be sufficient to capture the overall avalanche activity over a larger region. Intermediate resolutions may provide a reasonable compromise between the precision of individual outlines and large-area coverage at a reasonable cost and immediacy. Recording the perceived uncertainty while mapping might help (Hafner et al., 2022) as well as using the area of agreement from several mappings, or jointly discussing areas of disagreement. We generally recommend aiming for good illumination for mapping avalanches, especially under dry-snow conditions.

## 17.4 Limitations

In all three studies, we relied on comparably small sets of images: in study 1, the size survey, we aimed at a high response rate, which came at the cost of a smaller selection of different avalanche examples (10) since we did not want to introduce a bias by showing different images to different participants. The two mapping tasks (studies 2 and 3) were rather time-consuming and were, therefore, limited to few participants and to few examples. They may therefore be regarded as pilot studies, whose results should be interpreted with caution, keeping in mind the comparably small datasets and the potential particularities of the data. In particular, biases may be present due to the homogeneous sample from participants with a similar background and training, as well as due to the chosen oblique photos and remotely sensed data.

For study 1, we provided photographs but no maps and no additional information, for instance on damage which may have occurred. This certainly made the size estimation task somewhat more difficult, as often suspect either that a map is available or that the person is familiar with the avalanche path, which may both help in estimating avalanche dimensions. We did not provide maps together with the photographs, as we wanted to avoid introducing a bias related to the (un)familiarity with a specific map design. We have, however, tried to compensate the lack of an accompanying map through the presence of reference objects (people, trees, ski lifts, etc.) in our example photographs to help participants determine the potential damage and volume/mass. Nevertheless, photographs in Fig. 16.1a and e lack recognizable reference objects, and the other

photographs (Fig. 16.1) must be analyzed carefully to identify the ski tracks, people and houses. Agreement for avalanches next to clearly visible reference objects could therefore be higher than in our analysis.

For the avalanches mapped from oblique photographs (study 2), we speculate that study participants, being aware that their mappings will be analyzed in detail, may have paid more attention to finding the exact boundaries than during routine documentation work. We acknowledge that our sample size of six avalanche examples covers only a fraction of possible viewing angles, snow and avalanche conditions, and terrain characteristics. Furthermore, all participants were well acquainted with the study area and had extensive experience with mapping avalanches using oblique photos. Thus, we regard our results as a best-case scenario. Still, we believe that within the range of (fairly typical) conditions captured by our set of pictures, the evaluation is representative of avalanche outlines currently used in Switzerland. We encourage further research to ascertain the worldwide validity of the results.

Finally, we would like to point out that three of the authors were also involved as participants in the studies (one in study 1, and two in studies 2 and 3). Particularly in Studies 2 and 3, with few participants, this may impact results favorably and suggests the presented findings should be treated as an upper bound.

## Conclusions and Outlook

We quantified uncertainty related to avalanche size estimation and avalanche outline determination calculating the proportion of agreement between raters, the agreement with the most frequently chosen size and the agreement with the reference size. For avalanche outlines we investigated spatial agreement using the intersection over union between individual mappings as well as compared to a reference. Like in Van Coillie et al. (2014) the amount of variation depends on the type of task presented to the operator: we showed that it is difficult to consistently estimate avalanche sizes, and our analyses revealed significant differences between North American and European experts. The mapping of avalanches either from oblique photographs or from remotely sensed imagery proved to be a challenging task resulting in large intra-rater variabilities: some experts showed consistently larger deviations from the reference data. In the most extreme case this resulted in the deviation being 2 to 4 times larger than the smallest mapped avalanche area. For the mapping from remotely sensed imagery, individual experience and competence proved to have a larger impact than the underlying spatial resolution. Snow conditions also influenced agreement, with higher agreement under wet-snow compared to dry-snow conditions. For both mapping tasks our samples were fairly homogeneous and the sample size was rather small, limiting the generalizability of our results. Nevertheless, they shed light on the uncertainty underlying avalanche outlines for the first time. We strongly encourage further investigations into agreement between avalanche outlines to paint a more complete picture of the variation in the currently widely used and generated datasets.

Our findings indicate that the reliability of human estimates as a reference or ground truth for avalanche-related tasks needs to be questioned and critically assessed. Since these data are used as ground truth, for instance, for the validation of numerical avalanche simulations (e.g., Wever et al., 2018) or for training models estimating avalanche size from snowpack simulations (e.g., Mayer et al., 2023a), efforts should be made to obtain at least an approximate idea on the reliability of labels used when depending on them. Specifically efforts to average out unsystematic error and those requiring justification for the choice to endorse the analytic process (Stewart, 2001; Hagafors and Brehmer, 1983), may help to achieve more reliable results for the avalanche-related tasks presented in this paper. This could be achieved by relying on a superior method to obtain a ground truth or otherwise independent estimates of several experts to allow, for example, choosing the most frequent size.

Besides suggesting more precise definitions and training protocols, our results call for

automation. Modern image analysis algorithms, often based on machine learning (like Hafner et al., 2022, in the context of avalanche mapping from SPOT 6/7 – Satellite pour l’Observation de la Terre – imagery), are by no means perfect, but they rival human performance and offer consistent, repeatable results. Our reliability study may serve as a baseline to relate the outputs of such automatic methods to human expert performance. Even though the models cannot erase the inter-observer variability and will only learn to reproduce the outlines they are trained with, they can help to generate reproducible and comparable results.

# Part IV

## Interactive Snow Avalanche Segmentation from Webcam Imagery: results, potential and limitations

Elisabeth D. Hafner, Theodora Kontogianni, Rodrigo Caye Daudt,  
Lucien Oberson, Jan Dirk Wegner, Konrad Schindler, Yves Bühler

**The Cryosphere, published: 23.08.2024**

**Author version:** for the published typeset version please refer to Hafner et al. (2024)

**Author contributions:** EDH and TK came up with the initial idea, EDH coordinated the study, collected the images and annotated the avalanches used for training. EDH and LO adapted the model for avalanches and EDH did the analyses and organized the user study. TK, RCD, JDW and KS advised on the machine learning aspects of the project and critically reviewed the associated results. EDH wrote the initial paper and all the co-authors critically reviewed and complemented it.

## **Abstract**

For many safety-related applications such as hazard mapping or road management, well-documented avalanche events are crucial. Nowadays, despite the variety of research directions, the available data are mostly restricted to isolated locations where they are collected by observers in the field. Webcams are becoming more frequent in the Alps and beyond, capturing numerous avalanche-prone slopes. To complement the knowledge about avalanche occurrences, we propose making use of this webcam imagery for avalanche mapping. For humans, avalanches are relatively easy to identify, but the manual mapping of their outlines is time intensive. Therefore, we propose supporting the mapping of avalanches in images with a learned segmentation model. In interactive avalanche segmentation (IAS), a user collaborates with a deep-learning model to segment the avalanche outlines, taking advantage of human expert knowledge while keeping the effort low thanks to the model's ability to delineate avalanches. The human corrections to the segmentation in the form of positive clicks on the avalanche or negative clicks on the background result in avalanche outlines of good quality with little effort. Relying on IAS, we extract avalanches from the images in a flexible and efficient manner, resulting in a 90% time saving compared to conventional manual mapping. The images can be georeferenced with a mono-photogrammetry tool, allowing for exact geolocation of the avalanche outlines and subsequent use in geographical information systems (GISs). If a webcam is mounted in a stable position, the georeferencing can be re-used for all subsequent images. In this way, all avalanches mapped in images from a webcam can be imported into a designated database, making them available for the relevant safety-related applications. For imagery, we rely on current data and data archived from webcams that cover the Dischma Valley near Davos, Switzerland, and that have captured an image every 30 min during the daytime since the winter of 2019. Our model and the associated mapping pipeline represent an important step forward towards continuous and precise avalanche documentation, complementing existing databases and thereby providing a better base for safety-critical decisions and planning in avalanche-prone mountain regions.



## Introduction

Information on avalanche occurrences is crucial for many safety-related applications: for hazard mitigation, the dimensions of past avalanches are crucial for planning new and evaluating existing protection measures (e.g., Rudolf-Miklau et al., 2014). For the derivation of risk scenarios and the estimation of avalanche frequency, past events are an important piece of information as well (Bründl and Margreth, 2015). Mapped avalanches are also used to fine-tune and further develop numerical avalanche simulation software like SAMOS or RAMMS (Sampl and Zwinger, 2004; Christen et al., 2010). Today, information on past avalanches is still mainly reported and collected unsystematically from isolated locations, by observers and (local) avalanche warning services, though more recent research has proposed using satellite imagery (e.g., Eckerstorfer et al., 2016; Wesselink et al., 2017; Bianchi et al., 2021; Hafner et al., 2022). Depending on the source, these reports contain information on the avalanche type; the avalanche size; the approximate release time; the complete outlines, or at least the approximate location; the aspect; the type of trigger; and additional parameters. To enlarge the knowledge about avalanche occurrences, we propose a systematic recording of avalanches from webcam imagery. This usage of existing infrastructure allows for a large-scale application anywhere avalanche-prone slopes are already captured by webcams. Images can be acquired as frequently as needed without additional cost, enabling a near-real-time determination of release time. Furthermore, the sequence of images increases the chance of obtaining an image without low cloud cover or fog that would prevent documentation of the whole avalanche. Except for our own initial proposition (Hafner et al., 2023a) and that of Fox et al. (2023), we do not know of any attempt that makes use of this data source for avalanche identification and documentation. Fox et al. (2023) proposed two models in their initial experimental study for automatic avalanche detection from ground-based photographs: one for classifying images with and without avalanche occurrences and the other for segmenting the contained avalanches with bounding boxes. In opposition to their focus on finding the images and areas containing avalanches, we are aiming to extract the exact avalanche outlines from the imagery.

Detecting individual objects and determining their outlines is the objective of instance segmentation. This is important, for example, in the fields of autonomous driving (e.g., De Brabandere et al., 2017), remote sensing (e.g., Liu et al., 2022), and medical imaging (e.g., Chen et al., 2020). Numerous instance segmentation models have been proposed in recent years that are based on the superior image-understanding capabilities of deep learning. Besides

the quest for fully automatic methods, there is also an area of research dedicated to interactive object segmentation (IOS), where a human collaborates with the computer vision model to segment the desired object with high accuracy but low effort (Boykov and Jolly, 2001; Gulshan et al., 2010; Xu et al., 2016; Sofiiuk et al., 2020; Kontogianni et al., 2020; Lin et al., 2022; Kirillov et al., 2023). The human operator explicitly controls the segmentation, first by an initial input to mark the desired object (e.g., through a click or scribbles) and then by iteratively adding annotations to correct the segmentation where the automatic model makes mistakes, gradually refining the result. The goal is an accurate segmentation, provided by the IOS model with as little user input as possible. The key differences compared to instance segmentation are the user corrections and the way they are processed and encoded in the model. The vast majority of models proposed in recent years are employing clicks from the user for correcting the segmentation (e.g., Boykov and Jolly, 2001; Rother et al., 2004; Xu et al., 2016; Benenson et al., 2019; Kontogianni et al., 2020; Sofiiuk et al., 2021) and are using a combination of random sampling and simulating user clicks for training the model. The neighborhood of the clicked pixel is expanded to disks with radii of 3 to 5 pixels or to Gaussians, depending on the model. When disks are used to encode clicks, the whole area specified by the radius is given the same weight. When clicks are encoded as Gaussians, the weight is a Gaussian distribution, decreasing from the center of the click over the area specified by the radius.

There has been only little work on webcam (-like) imagery; the dominant data source for automatic avalanche documentation so far has been satellite imagery (e.g., Bühler et al., 2019; Eckerstorfer et al., 2019; Hafner et al., 2021; Bianchi et al., 2021; Karas et al., 2022; Kapper et al., 2023). Optical satellite data, proven to be suitable to reliably capture avalanches (spatial resolution of approx. 2 m or finer; Hafner et al., 2021; Hafner, 2023), need to be ordered and captured upon request which is expensive and dependent on cloud-free weather conditions. Radar data have the big advantage of being weather independent, but with one satellite in operation, open-access Sentinel-1 data is only available at selected dates (currently, approximately every 12 d in Switzerland) and other suitable radar data need to be ordered and purchased as well. Additionally, with a spatial resolution of approximately 10–15 m, it is not possible to confidently map avalanches of size 3 and smaller from Sentinel-1 imagery (Hafner et al., 2021; Keskinen et al., 2022). Furthermore, the exact or even approximate time of avalanche release cannot be retrieved from satellite data and remains unknown. However, where suitable satellite data are available, areas affected by avalanches may be identified and documented continuously over large regions with identical methodology.

Applications relying on information about avalanche occurrences not only seek confirmation of an avalanche near a specific webcam but also require details such as the precise location, extent, aspect of the release area, and size of the avalanche. Avalanches captured on oblique photographs may be georeferenced to enable a transfer of the avalanche identified in the image to a map. There are several mono-photogrammetry tools available to georeference single images, initially developed to georeference historic photographs (e.g., Bozzini et al., 2012, 2013; Produit et al., 2016; Golparvar and Wang, 2021). Only with existing georeferencing can the detected avalanches be exactly geolocated; compared by size, aspect, or slope angle; and imported into existing long-term databases. Since most webcams are mounted in a stable position, always capturing the same area, the georeferencing only needs to be done once and may be re-used for all subsequent images.

To complement the currently established ways avalanche occurrences are documented, we propose to use webcam infrastructure, regularly acquiring imagery for avalanche mapping. In the present work, we identify avalanches in imagery employing interactive object segmentation

(interactive avalanche segmentation, IAS). Since human user interactions are modeled during training we investigate the transferability of our model results to the real-world use by humans in a user study. We use webcam imagery from stations maintained by the WSL Institute for Snow and Avalanche Research SLF (SLF) available every 30 min in near-real-time, and the avalanche library published by Fox et al. (2023). Additionally, we propose a workflow to georeference the identified avalanches with the mono-photogrammetry tool from Bozzini et al. (2012, 2013). By mapping avalanches from webcam imagery, we enlarge existing avalanche databases, thereby allowing for better decision-making for downstream applications.

## 20.1 SLF Webcam network

Our webcam network covers the Dischma Valley, a high alpine side valley of Davos, with 14 cameras mounted at six different locations (Fig. 20.1). The valley is about 13 km long, and the valley floor reaches from 1500 m a.s.l to 2000 m a.s.l, while the summits reach heights over 3000 m a.s.l. The Dischma Valley is permanently inhabited in the lower 5 km while the road leading to its upper part is closed in winter. Steep mountains are located on both sides of the valley, and over 80% of the entire area constitutes potential avalanche terrain (Bühler et al., 2022). Outside the permanent settlements, avalanches can only be monitored remotely, especially during high avalanche danger.

Each of our six stations is equipped with two to three cameras (usually a Canon EOS M100), operated with an independent power supply with a solar panel and a battery, except for Stillberg, where we connected to existing power lines (Fig. 20.2). The acquisition of images every 30 min during daylight is programmed and automatically triggered by a small on-station computer. This interval lowers the risk of cloud cover and captures avalanches under different illumination conditions once they have occurred. The images are then sent to SLF in near-real-time via the mobile network and are stored on a server. The first camera was mounted at the Büelenberg station in the summer of 2019, with the next four stations being established in the following months. The Börterhorn station came later, having only been in operation from December 2021 to June 2023, and was moved to a new location with a similar view in December 2023 (Hürel station). The images have previously been used in the ESA DeFROST project (ESA, 2020) and in Baumer et al. (2023).

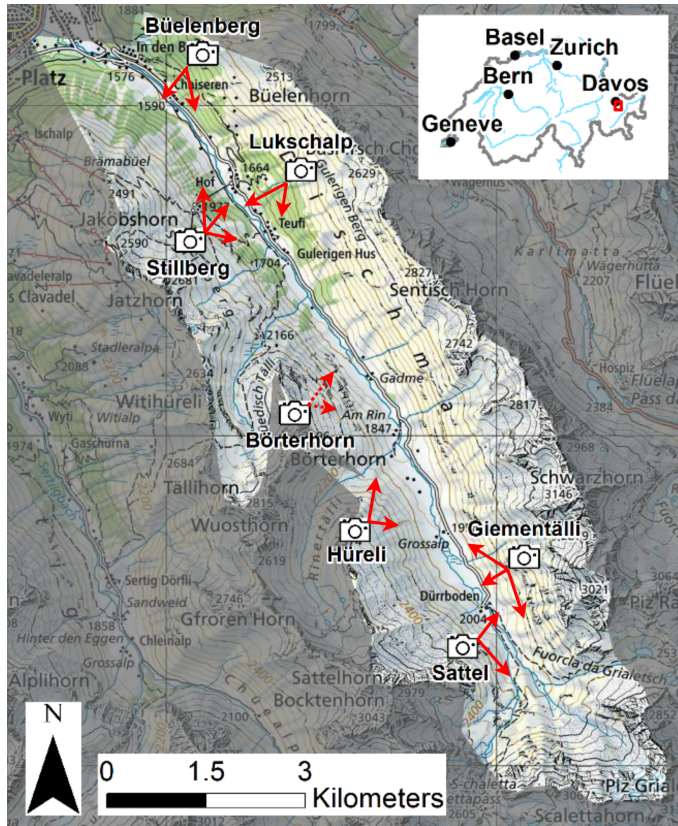


Figure 20.1: Locations, view directions (red arrows), and area covered by the 14 cameras mounted in six different locations in the Dischma Valley, Davos. The Hürel station succeeded the Börterhorn station (dashed arrows), which is no longer in operation (map source: Federal Office of Topography).



Figure 20.2: The stations in the Dischma Valley were either bolted to rock faces (left; Lukschalp) or mounted on a mast (right; Sattel). They host two to three cameras and all infrastructure necessary to ensure power supply as well as data acquisition and transmission.

## 20.2 Avalanche images and annotations

We used unique sets of images for the model to learn from (training), for the unbiased evaluation during training and hyperparameter tuning (validation), and for the unbiased evaluation of the final model (testing).

### 20.2.1 SLF dataset

We rely on imagery from the webcams at our stations for training (all except Börterhorn and Hüreli; Sect. 20.1), validation, and testing. The images, with a size of  $6000 \times 4000$  pixels, are from seven different cameras that have captured identifiable avalanches well since being in operation. For training, we prepared the images and cropped them to  $1000 \times 1000$  pixels, keeping only the avalanches and their immediate surroundings in the original resolution. For evaluation and for our user study, we want to segment all captured avalanches per image; therefore we only resized the images to  $3600 \times 2400$ , the largest the model may handle.

The avalanches in the images were manually annotated with the smart labeling interface provided by Supervisely (Supervisely, 2023). The SLF dataset contains roughly 400 annotated avalanches (Tab. 20.1). About three-quarters of the avalanches are used for training, testing, and validation, while the rest are used to test generalizability. For this, we use images with a certain domain gap relative to the training images: 46 images from the two Börterhorn webcams, excluded from training (WebNew) and a set of 44 images taken from handheld cameras (GroundPic; Tab. 20.1). The WebNew contains mostly small avalanches, some of them captured under diffuse illumination conditions, while the GroundPic depicts larger avalanches and includes some lower quality images taken with mobile phones. For our user study, we relied on a combination of different webcam images showing avalanches of different sizes and captured under varying illumination conditions. Of the 20 annotated avalanches (UserPic), 75% are unique to the dataset, while the rest are also part of the WebNew or the GroundPic.

Table 20.1: Overview of the datasets used.

Dataset name	Avalanche an-notations	Description	
SLF	train	200	Webcam imagery and annotations from our test site in Dischma (Fig. 20.1).
	vali	44	
	test	45	
	WebNew	46	Imagery and annotations from the Börterhorn station (Fig. 20.1), whose two webcams were excluded from the SLF train, vali and test and have an unseen viewpoint relative to these images.
	GroundPic	45	Imagery and annotations taken from handheld cameras with an unseen viewpoint relative to all training images.
	UserPic	20	Imagery from webcams and corresponding annotations. 75% of the images are unique to this dataset while the rest are also part of the WebNew or GroundPic.
UIBK	train	2102	Imagery and annotations used by Fox et al. (2023).
	vali	382	
	test	867	

## 20.2.2 UIBK dataset

Fox et al. (2023) have published a dataset containing images of over 3000 avalanches from different perspectives with annotations of the avalanche type (slab, loose snow and glide snow avalanches; University of Innsbruck et al., 2023). In addition to avalanches, their category “glide snow avalanche” also contains glide snow cracks where no avalanche has occurred (yet). We decided to include a selection of their annotations in some of our training configurations to evaluate the performance of our setup using a multi-source dataset. We are, however, interested in avalanches only; therefore, we manually sorted out images with glide snow cracks and excluded them for training. Consequently, we used a subset of 2102 binary avalanche masks from the UIBK dataset for training and 382 avalanches for validation, which we prepared by cropping to  $1000 \times 1000$  pixels (Tab. 20.1). For the test dataset, we kept all images, depicting 867 avalanches and glide snow cracks, to allow for a fair comparison to Fox et al. (2023). Fox et al. (2023) provide no details about the manual annotation procedure. We note that upon comparison, their annotations are markedly coarser than ours, with significantly smoother and more generalized avalanche outlines (e.g., Fig. 20.3). We resized the images larger than  $3600 \times 2400$  to that size for the evaluation.

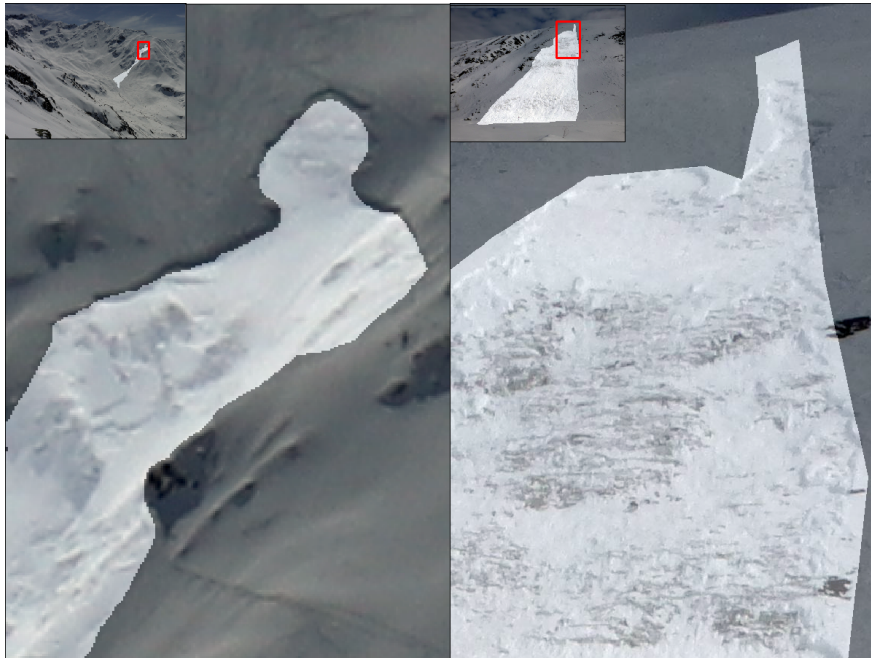


Figure 20.3: Comparing the details in the annotation from one of the SLF webcam images (left) to an image from the UIBK dataset (right; University of Innsbruck et al. (2023)).

## Methodology

We used a state-of-the-art interactive image segmentation model (Sofiiuk et al., 2021), modified it for avalanches, and trained it with three different sets of avalanche imagery. The trained model was then applied to new, unseen images to qualitatively and quantitatively evaluate the resulting avalanche outlines using both per-pixel and per-avalanche metrics. It is important to note that click locations have to be selected automatically to enable large-scale training and testing. This could lead to performance differences caused by deviations between simulated clicks and real user behavior. We therefore additionally designed and carried out a user study with human annotators to ascertain that the efficiency gains carry over to the real use case.

### 21.1 Model architecture

We employed the interactive segmentation model introduced by Sofiiuk et al. (2021), adapted it specifically to avalanches, and trained it with a variety of avalanche datasets. Sofiiuk et al. (2021) used the HRNet+OCR method, a high-resolution network (HRNet) with an added object-contextual-representation (OCR) module (Wang et al., 2020; Yuan et al., 2020; Xu and Zhao, 2024). The HRNet+OCR architecture connects high- and low-resolution convolutional processing streams in parallel and enables information exchange across different resolutions (Wang et al., 2020). The OCR module explicitly accounts for the global context to achieve better segmentation of objects in complex images (Xu and Zhao, 2024), which is particularly valuable in our case, where avalanches that can make up large parts of the images while being hard to distinguish from the white snow in the background if considering only local evidence. Positive and negative click locations from interactive user input were encoded as disks with a fixed radius of 5 pixels (Benenson et al., 2019).

Semantic segmentation backbones usually take only RGB images as input; for interactive segmentation, the handling of additional model input in our case, encoded user clicks needs to be carefully implemented (Fig. 21.1). Sofiiuk et al. (2021)'s solution to this is Conv1S: a convolutional block that outputs a tensor of exactly the same shape as the first convolutional block in the backbone. The output of the first backbone convolutional layer (usually 64 channels) is then summed up element-wise, with the convolutional block being applied to the encoded user clicks. With this implementation, it is possible to choose a different learning rate for new weights without affecting the weights of a pre-trained backbone.



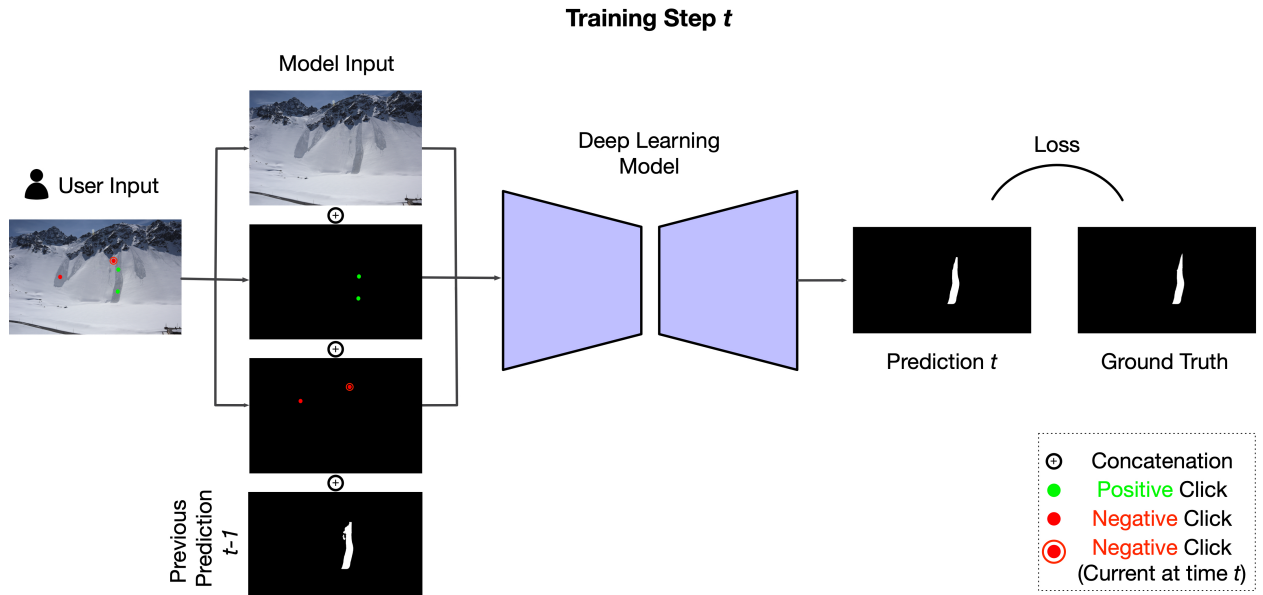


Figure 21.1: Illustration of the fine-tuning step of the IOS when training on avalanches.

A combination of random and iterative sampling strategies are employed to simulate human user clicks for training, with masks from previous steps being included in the iterative sampling procedure (Fig. 21.2). Morphological erosion is used to shrink the largest mislabeled region before setting the sampling point into its center, which proved to be superior to simply setting the next click in the center of the erroneous region (Mahadevan et al., 2018). The click may be positive, denoting the avalanche, or negative for the background. In the evaluation mode, the click is put at the center of the largest erroneous region, be it false positive or false negative, as proposed in Xu et al. (2016) and Li et al. (2018). The maximum number of clicks (positive or negative) is set to 20 for both training and evaluation.

We made the following adaptations to the original model from Sofiiuk et al. (2021):

- We trained on patches of  $600 \times 600$  pixels instead of patches of  $320 \times 480$  pixels, which we cropped from varying places of our training images.
- For data augmentation during training, we additionally included random translation (max. 3%) and rotation (max.  $10^\circ$ ).
- We replaced the manual multistep learning-rate scheduler with a cosine learning-rate scheduler to profit from a decreasing learning rate without the need to fiddle with the steps and rates of decay.
- We did not use the zoom-in function.
- We used a batch size of 4 instead of 28 due to our relatively small training dataset but fine image resolution.

## 21.2 Evaluation metrics

The raw predictions (i.e., the per-pixel probabilities for being part of the avalanche) were thresholded at 0.5 to obtain binary avalanche masks for the analyses. We used the intersection over union (IoU) as an indicator of spatial agreement between either the predicted and ground truth masks or the bounding boxes around those masks (e.g., Levandowsky and Winter, 1971).

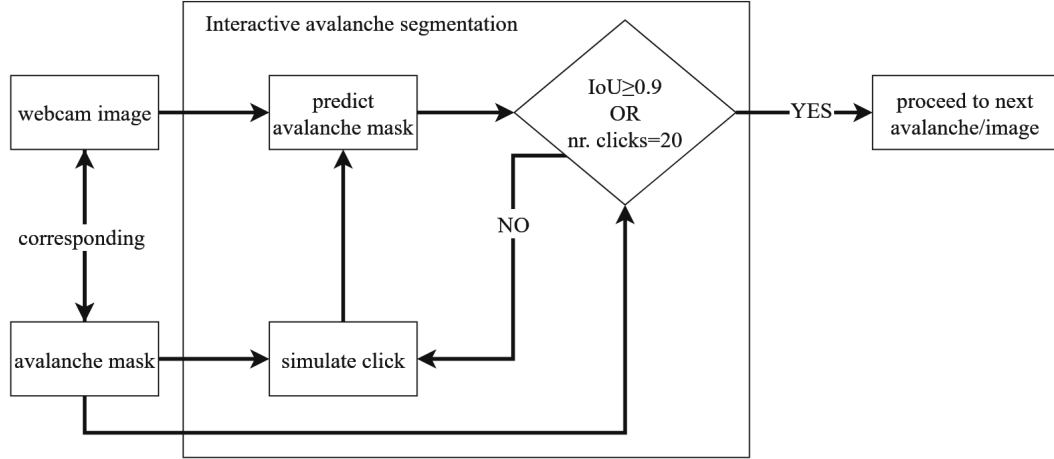


Figure 21.2: Illustration of the handling of one avalanche when training the IAS model with clicks generated by random and iterative sampling. For the new prediction, all previous clicks, as well as the previous mask (if available), are considered.

### 21.2.1 Pixel-wise metrics

On the pixel level of the masks, we recorded the average number of clicks (NoC) necessary to reach IoU thresholds of 0.8 and 0.9 (denoted as  $mNoC@80$  and  $mNoC@90$ , respectively). Achieving a high IoU after few clicks makes the model most useful. Consequently, we compared the IoU at click  $k$  (for  $k = 1, 2, \dots, 20$ ) averaged over all the images ( $mIoU@k$ ). Additionally, we calculated the number of images that do not reach 0.85 IoU, even after 20 clicks ( $NoC_{20}@85$ ).

### 21.2.2 Object-wise metrics

On the object level, we compared the IoU of the bounding box of the predicted and that of the ground truth avalanche annotation. If the IoU between two bounding boxes is larger than or equal to a threshold  $T$ , the detection is considered to be correct, while, for values below the threshold  $T$ , it is not (Padilla et al., 2020). Like Fox et al. (2023), we first considered a  $T \geq 5\%$  between the bounding boxes to be a match, but, additionally, we evaluated with  $T \geq 50\%$ , which is a more standard value in literature (Redmon et al., 2016; He et al., 2018).

From the matches, we computed the F1 score as

$$F1 = 2 \cdot \frac{PPV \cdot POD}{PPV + POD}, \quad (21.1)$$

where the probability of detection (POD) and positive predictive value (PPV) are defined as

$$POD = \frac{TP}{TP + FN} \quad \text{and} \quad PPV = \frac{TP}{TP + FP}, \quad (21.2)$$

where TP is true positive, FP is false positive and FN is false negative.

### 21.2.3 Comparison of time needed

The time spent to map an avalanche with the “traditional method”, such as with the avalanches that are part of the DAvalMap inventory (Hafner et al., 2021), is not recorded by default. For a comparison to the time spent on IAS, we had one experienced person record the number of minutes needed for manually mapping 274 avalanches from photographs (mean size of 1.75; European avalanche size definition (EAWS, 2023)) with the methodology described in Hafner (study 2; 2023).

## 21.3 Experimental setup

To find the best model for interactively segmenting avalanches from our webcam imagery, we evaluated several training regimes, all with the same model architecture but varying training datasets (see Sect. 21.1). Our baseline was the model trained only on COCO+LVIS (104 000 images and 1.6 million instance-level masks; Lin et al., 2015; Gupta et al., 2019), meaning that it has never seen an avalanche. We then trained three further versions, re-using the knowledge already learned from being trained on COCO+LVIS, and fine-tuning the model with different sets of avalanche data: AvaWeb, trained on the SLF dataset, AvaPic, trained on the UIBK dataset; and AvaMix, trained on a combination of those two (Tab. 20.1). Preliminary tests confirmed that fine-tuning the model pre-trained on COCO+LVIS was always superior to training from scratch using only avalanche data. This is in line with previous work on avalanches (Hafner et al., 2022). We performed hyperparameter tuning on the validation set (e.g., selecting the ideal number of training epochs 90 for AvaWeb and AvaPic and 95 for AvaMix – using a threshold of 0.5 on the raw predictions). We used the hyperparameters selected for the validation set, fixed during our evaluation, of the test set. For the evaluation, we checked how well the model generalizes to the SLF test as well as to images from other webcams (WebNew). We additionally evaluated the GroundPic and the UIBK test to assess the robustness of the model configurations in relation to images from outside our webcam perspective. In addition, we compared to segmentation results from previous work by Fox et al. (2023) by calculating bounding boxes for our predictions and evaluating their overlap with respect to the ground truth bounding boxes from the UIBK test.

## 21.4 User Study

We carried out a small user study to investigate if the metrics from evaluating our model hold with real users, whose inputs are noisier and who may adapt to model behavior. Eight participants were given a short introduction and mapped one avalanche per UserPic image. For our user study, we used the graphical user interface (GUI) provided by Sofiiuk et al. (2021), adapting it to save the click coordinates, the time needed per click, and the predicted masks for each click together with the IoU. Since several images captured more than one avalanche, we added an arrow pointing at the desired avalanche in each UserPic image. Before segmenting the marked avalanches in UserPic, the participants performed two trial segmentations that were not used for evaluation to familiarize themselves with the GUI, the annotation protocol, and the data characteristics. Participants were allowed a maximum number of 20 clicks per avalanche but were told they could stop earlier if they were satisfied with the segmentation. As metrics for the user study, we calculated the mNoC@80 and mNoC@90 and compared the mIoU@k, the mean annotation time, the NoC<sub>20</sub>@85, and the differences between the best and worst results in terms of mean IoU. To investigate variability in the avalanche areas identified, like in Hafner (2023), we calculated pairwise IoU scores for the final masks based on the last employed click per participant. To test whether the differences between the mIoU scores of the participants are statistically significant, we used the two-sided  $t$  test (as implemented in R Core Team, 2021), with a significance level  $p \leq 0.05$ .

## Results

### 22.1 Pixel-wise metrics

Evaluating based on the SLF test, the model trained on the AvaWeb was almost 10% better than the others and almost 25% better than the baseline (COCO+LVIS; Fig. 22.1) from click 1. It remained on top, but the others caught up by approximately click 16. AvaPic was consistently the worst at high click numbers and even dropped below the baseline. Adding the data from AvaWeb to AvaPic in the AvaMix improved the results, but the quality reached only half that of the results of the AvaWeb alone. Compared to the baseline, all models trained with avalanches were superior to the baseline, especially for the first half of the clicks and, except for the AvaPic, for the last half of the clicks. Overall, the AvaWeb needed the least clicks to reach the desired IoU thresholds, and it was only for one image that it never reached the  $\text{NoC}_{20}@85$ . The AvaPic never reached this threshold for five images, while this was the case for only two images for the AvaMix, and even the baseline reached an IoU of 85% for more images. For the remaining analyses, we did not consider the model trained only on COCO+LVIS (baseline).

Table 22.1: Results for the different datasets when evaluating based on the SLF test. The bold values highlight the best metrics.

Model	Pretrained weights	mIoU@1 (%)	mIoU@2 (%)	mIoU@3 (%)	mNoC @80	mNoC @90	NoC <sub>20</sub> @85
COCO+LVIS (baseline)	—	35.07	52.62	65.00	5.58	9.42	3
AvaWeb	COCO+LVIS	<b>58.59</b>	<b>73.40</b>	<b>78.30</b>	<b>3.31</b>	<b>7.60</b>	<b>1</b>
AvaPic	COCO+LVIS	48.50	62.51	69.42	5.24	10.73	5
AvaMix	COCO+LVIS	49.75	66.24	73.03	4.11	9.40	2

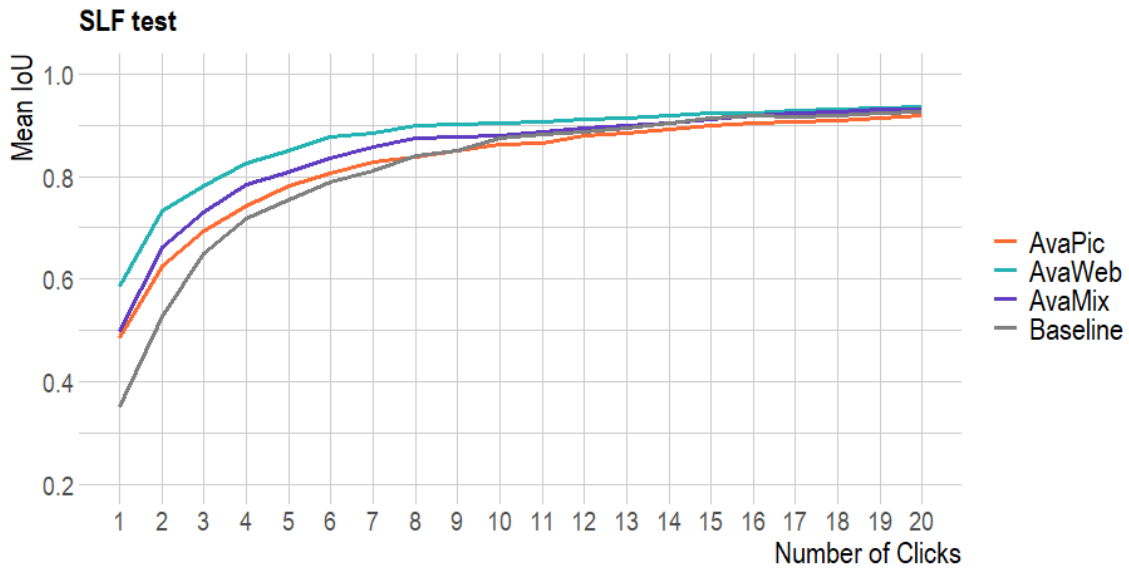


Figure 22.1: Evaluation on the SLF test based on models trained with different datasets: baseline (COCO-LVIS), AvaWeb (SLF train), AvaPic (UIBK train), and AvaMix (SLF + UIBK train).

To check how well the models generalize to new avalanches under varying perspectives, we evaluated them based on the WebNew, the GroundPic and the UIBK test (Fig. 22.3; Tab. 22.2): AvaWeb was superior, with a margin of up to 30% from click 1 over the AvaPic and AvaMix based on the WebNew (Tab. 22.1; Fig. 22.2). The AvaPic and AvaMix only caught up around click 10 but never surpassed the AvaWeb. For all models, the images in the NoC<sub>20</sub>@85 category depicted small, often long and slim avalanches located in the shade in imagery acquired under diffuse-illumination conditions and/or avalanches that had been snowed on, reducing the overall visibility of the features (Fig. 22.4).

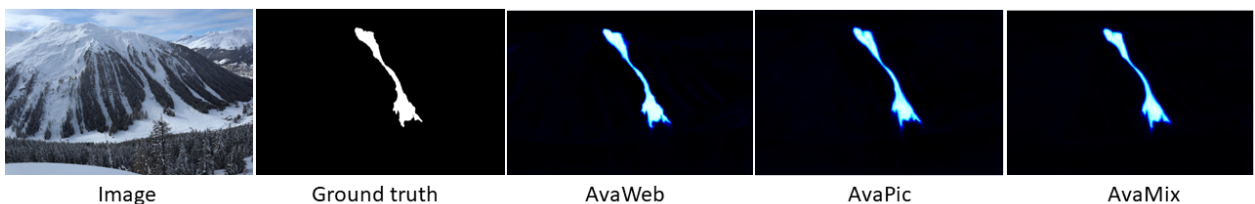


Figure 22.2: Example of an image from the SLF test that all three models solve well. The lighter the hue in the model predictions, the higher the model certainty concerning the existence of an avalanche. In a close-up look, the AvaWeb prediction exhibits more nuanced and detailed avalanche boundaries.

Based on the ground-based GroundPic, the AvaWeb started out being the worst by a margin of about 10%, but it caught up and surpassed the AvaPic from click 5 onwards, though it never reached the AvaMix. For the large but more coarsely annotated UIBK test, the AvaPic and the AvaMix were consistently superior to the AvaWeb by 10 to 20%. The AvaWeb struggled the most with ground-based close-up views of avalanches, often in combination with diffuse-illumination conditions or shade, as well as avalanches captured on coarse images from mobile phones (Fig. 22.5). For some of those avalanches, the IoU score reached after 20 clicks is well below 50%. For more than one-quarter of all avalanches, the AvaWeb never reached the NoC<sub>20</sub>@85,

while, for the AvaPic and AvaMix, less than 1% of all avalanches never reached an IoU of 85%. The AvaPic and AvaMix struggled mostly with the same images, which depicted close-up views of the release area of avalanches in diffuse illumination conditions or avalanches which have been snowed on and are hard to spot.

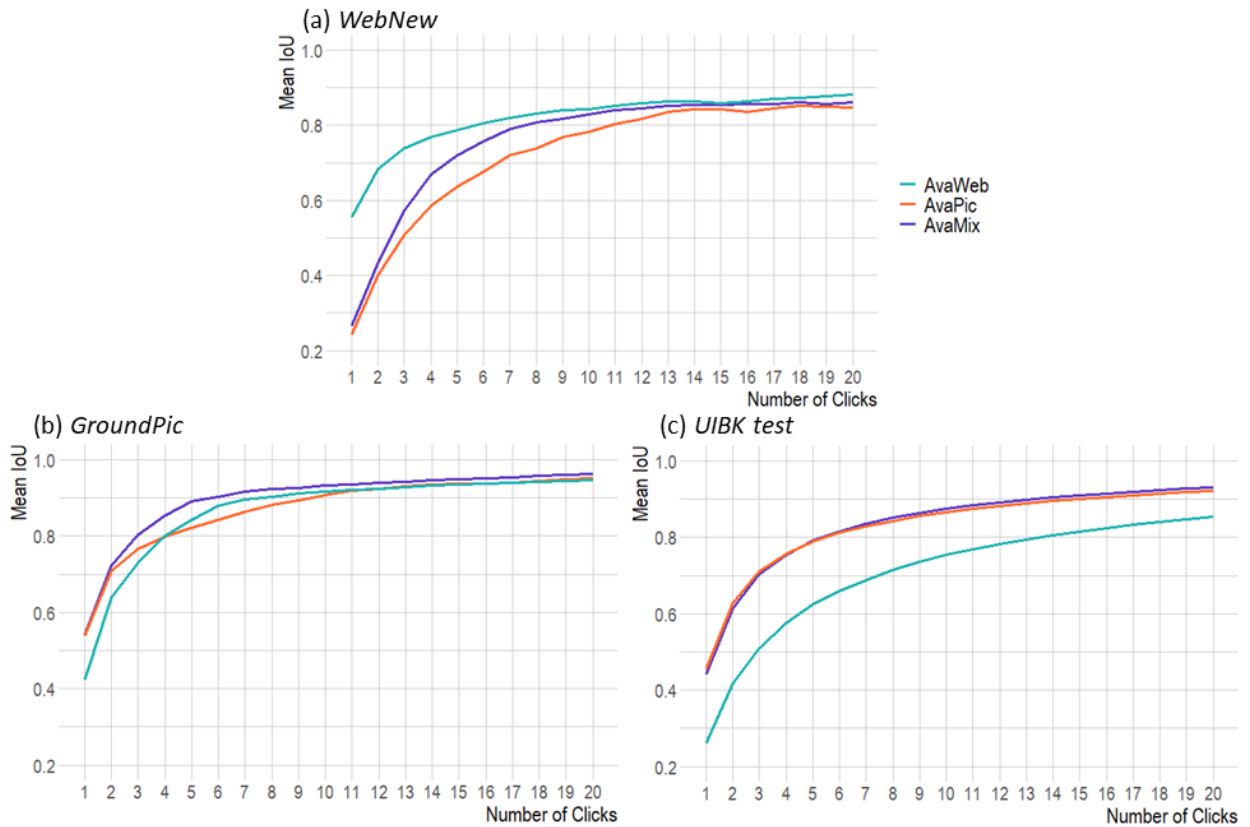


Figure 22.3: Comparing mIoU per click for three datasets with a domain gap in relation to the initial webcam data for our three training configurations: AvaWeb (SLF train), AvaPic (UIBK train) and AvaMix (SLF + UIBK train).

Table 22.2: Results when evaluating the generalizability in relation to data not seen during training, with a domain gap with respect to the training data. The bold values highlight the best metrics.

Dataset	Model	mIoU@1 (%)	mIoU@2 (%)	mIoU@3 (%)	mNoC @80	mNoC @90	NoC <sub>20</sub> @85
WebNew	AvaWeb	<b>55.61</b>	<b>68.24</b>	<b>73.85</b>	<b>6.65</b>	<b>13.57</b>	<b>12</b>
	AvaPic	24.31	40.08	50.76	10.78	16.07	15
	AvaMix	26.72	43.26	57.20	9.07	14.39	14
GroundTest	AvaWeb	43.32	63.43	73.38	4.53	<b>6.91</b>	2
	AvaPic	54.63	71.25	76.92	3.98	7.73	2
	AvaMix	<b>54.82</b>	<b>72.72</b>	<b>80.51</b>	<b>3.09</b>	6.96	<b>1</b>
UIBK test	AvaWeb	26.19	41.71	51.05	10.47	15.82	246
	AvaPic	44.28	61.29	70.37	<b>5.84</b>	<b>11.26</b>	<b>50</b>
	AvaMix	<b>45.70</b>	<b>62.67</b>	<b>70.99</b>	6.06	11.72	75



Figure 22.4: Example of an image from the WebNew with diffuse illumination and a long and slim avalanche that all three models struggled with. The lighter the hue in the model predictions, the higher the model certainty concerning the existence of an avalanche.

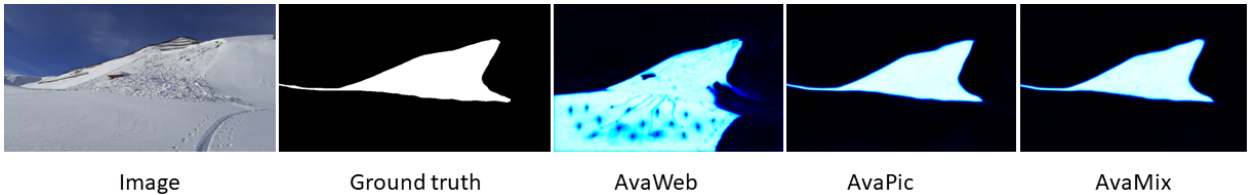


Figure 22.5: Example of a close-up view of an avalanche from the GroundPic, where the AvaWeb struggled with correctly identifying the avalanche area close to the photographer. The lighter the hue in the model predictions, the higher the model certainty concerning the existence of an avalanche.

## 22.2 Object-wise metrics

Comparing bounding boxes, the AvaWeb achieved an F1 score 0.12 points higher than that of Fox et al. (2023) from the first click onwards (0.64 vs. 0.76; bounding-box threshold of 0.05; Tab. 22.3). For both the AvaPic and the AvaMix, the F1 score was even close to 1, therefore being superior to Fox et al. (2023) by 0.33 to 0.34 and higher than the AvaWeb. With a threshold of 0.5 for the overlap of the bounding boxes, the scores were lower and lay between 0.23 (AvaWeb) and 0.44 (AvaPic) for the first click. Consequently, the AvaPic and the AvaMix were again superior to the AvaWeb (by around 0.2) and also remained on top for click 3 and 5. For click 5, the AvaPic and the AvaMix already achieved an exceptionally good F1 score above or equal to 0.94. No comparison to Fox et al. (2023) was possible for the 0.5 bounding-box threshold.

Table 22.3: Comparison of F1 scores and standard deviations (SDs) for the two different IoU thresholds (5% like Fox et al. (2023) and 50%) for the UIBK test. The bold values highlight the best metrics.

F1 score $\pm$ SD	Fox et al. (2023)	AvaWeb	AvaPic	AvaMix
		<b>IoU 5%</b>		
Automated	0.64 $\pm$ 0.60	—	—	—
Click 1	—	0.76 $\pm$ 0.43	<b>0.97 <math>\pm</math> 0.16</b>	0.96 $\pm$ 0.20
Click 3	—	0.99 $\pm$ 0.11	<b>1</b>	1 $\pm$ 0.05
Click 5	—	1 $\pm$ 0.08	<b>1</b>	<b>1</b>
		<b>IoU 50%</b>		
Click 1	—	0.23 $\pm$ 0.42	0.44 $\pm$ 0.50	0.42 $\pm$ 0.49
Click 3	—	0.66 $\pm$ 0.47	0.86 $\pm$ 0.34	0.87 $\pm$ 0.31
Click 5	—	0.80 $\pm$ 0.40	0.94 $\pm$ 0.24	0.96 $\pm$ 0.20

## 22.3 User study and time saved

For our user study, we loaded the model trained on AvaWeb for making predictions upon user input. On average, the participants employed 4.9 clicks for the UserPic, with variations from 1.25 to 9.63 clicks for the 20 different images. The employed clicks were on avalanches in 79% of all cases, while the rest were on the background. The avalanches that needed fewer clicks to reach a certain IoU threshold tended to be the smaller ones. Even though not everyone always clicked until an IoU of 85% was reached, on average, only one image remained below that value. This image depicted an avalanche that was located in a partly shaded and partly illuminated area, where, especially in the shade, features are hard to identify. On average, participants needed 6.5 s to reach an IoU of 80% and 9.1 s to reach an IoU of 90%. In opposition, on average, 2 min and 36 s were required for mapping one avalanche with the traditional method, with time needed ranging from 1 to 8 min. This is more than 2 min extra than when relying on IAS and translates to a saving in time of more than 90% compared to a manual mapping.

In our user study, we observed large variations between the different participants in terms of the average number of clicks (2.90 to 8.10), the mNoC@80 (1.80 to 2.80), and the mNoC@90 (2.00 to 3.12). Additionally, for avalanches such as that in Fig. 22.6 (top), there was no clear “middle” on which to place the first click, which resulted in very diverse click strategies for the participants. In contrast, for the avalanche in Fig. 22.6 (bottom), clicks were placed more homogeneously: first in the middle and then at the top and bottom, thereby correcting details. For clicks 1 to 5, where we had enough samples from all participants, we checked if the differences between the highest and the lowest mIoU values were statistically significant. The differences were not significant for



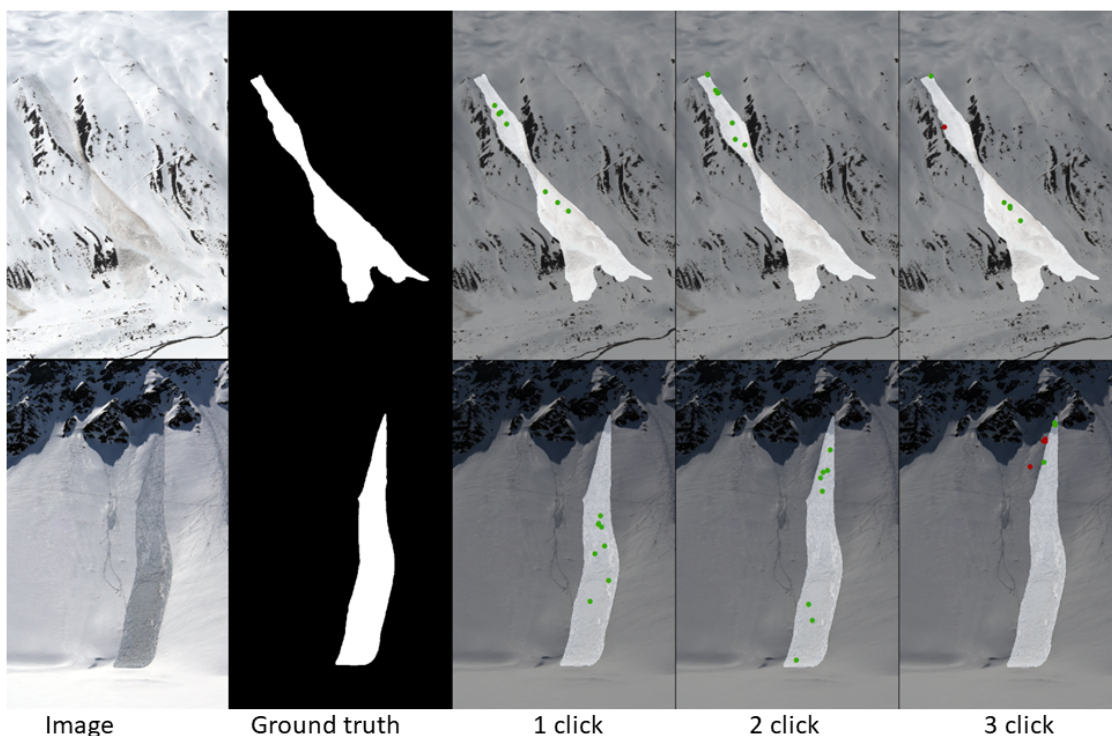


Figure 22.6: Illustration of where the first three clicks in two images from the UserPic dataset were placed. Green dots denote positive clicks, while red dots denote negative clicks.

IoU@1 and IoU@2 (t-test;  $p$  value  $> 0.05$ ) but they were statistically significant for IoU@3 ( $p$  value= 0.045), IoU@4 ( $p$  value= 0.034), and IoU@5 ( $p$  value= 0.035). This was caused by very consistent results with low standard deviations for the participants with the highest mIoU@k scores. However, when taking the mask from the last click as a final result, the differences between participants were quite small: the mean pairwise IoU was 93.53%, the maximum was 95.44%, and the minimum was 90.59%. Consequently, all pairs had an IoU within 5% of each other as their segmented final avalanche masks were very similar (Fig. 22.8).

When evaluating the model trained on AvaWeb based on the UserPic with simulated clicks and comparing to the user study results (see Tab. 22.4), the AvaWeb results were superior for all investigated metrics, except for the mNoC@80. The participants with the highest mIoU@k held up in comparison to the numbers from the model (Fig. 22.8).

Table 22.4: Comparison of the results from the user study with the model results when evaluating on the same imagery (UserPic; N = 20).

	User Study	AvaWeb
mNoC@80	2.11	1.85
mNoC@90	2.50	2.55
NoC <sub>20</sub> @85	1	0
mIoU@1 (%)	66.61	74.31
mIoU@2 (%)	80.91	89.57
mIoU@3 (%)	86.22	91.53

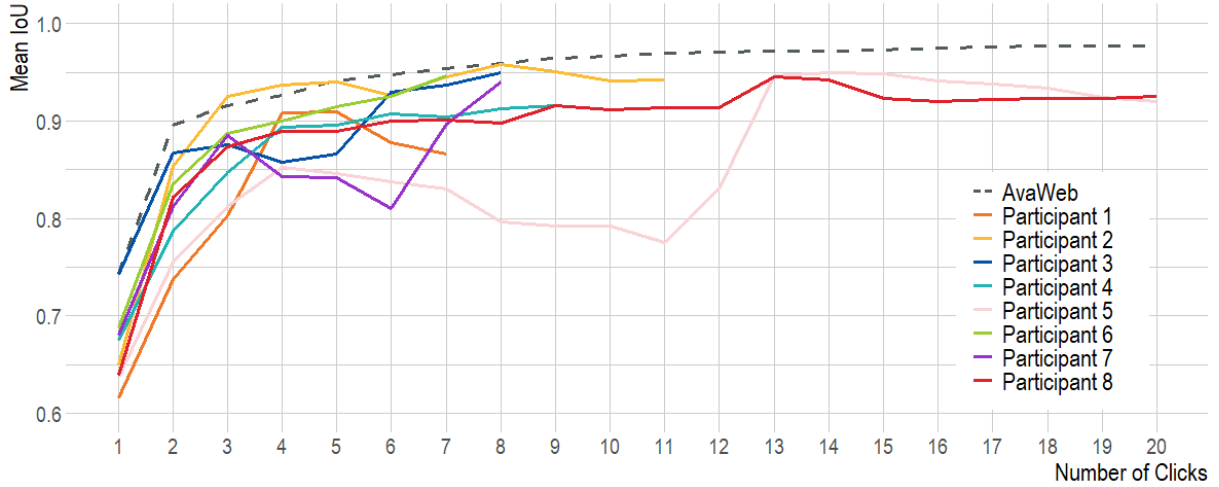


Figure 22.7: Comparison of the mIoU for all participants of the user study to the mIoU of the AvaWeb evaluated based on the UserPic dataset. Note that only two participants used the maximum possible number of 20 clicks.

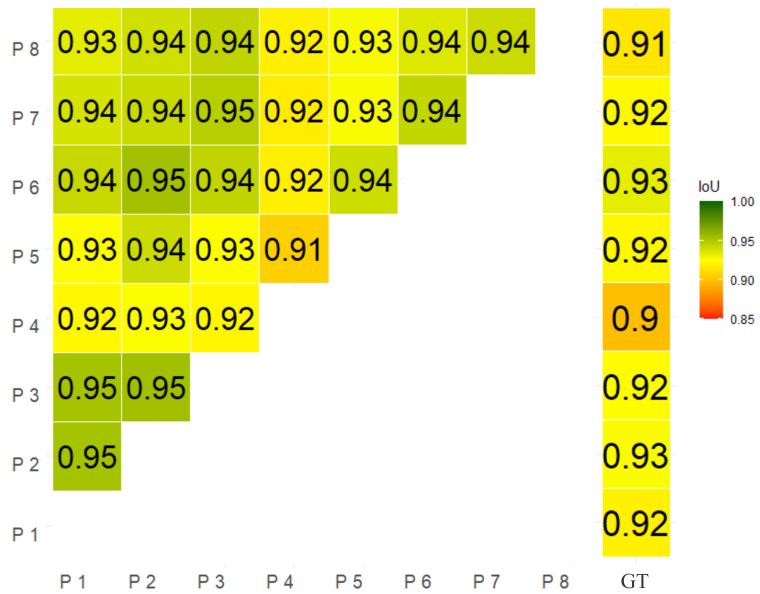


Figure 22.8: IoU for all participant pairs (participants denoted as P, and the ground truth as GT) for the final masks from our user study on the UserPic.

## Discussion

Our results show that IAS enables the segmentation of avalanches from webcam imagery within seconds. We compared the performance of the model trained with different datasets: as expected the model trained with any avalanche dataset outperformed the baseline (COCO+LVIS). The model trained on AvaWeb performed best for the two test datasets containing webcam imagery (SLF test and WebNew) and performed on par with the dataset with a perspective unlike those of the webcams (GroundPic) but failed to generalize well to the large but coarsely annotated UIBK test with a large variety of perspectives and resolutions. In contrast, the model trained on larger and more diverse datasets (AvaPic and AvaMix) exhibited lower mIoU scores and a higher number of clicks to reach a certain IoU for all test sets containing webcam imagery (SLF test and WebNew), but it performed better with imagery not from webcams (GroundPic and UIBK test). The model trained on AvaMix seems to have learned more details since the mIoU scores were higher than for the AvaPic for three out of four datasets from approximately click 3 to 10. During those clicks, after the initial coarse segmentation, details of the avalanche are segmented. We suspect that the detailed annotations, following the visible texture from the SLF dataset, helped the AvaMix to outperform the AvaPic.

Overall, the model struggled with images of avalanches recorded under unfavorable illumination conditions. This is in line with previous studies that found the agreement between different experts for manual mapping to be lower in shaded areas (Hafner et al., 2022; Hafner, 2023). Furthermore, in particular the AvaWeb struggled with close-up views of avalanches; often, these images are photographed from below the avalanche, resulting in a very specific perspective that the model has never seen during training. But overall, the AvaWeb, with less than 10% of the training data of the other two datasets, achieved the best performance for two out of three test sets with detailed avalanche annotations (SLF test, WebNew, GroundPic). Even though the UIBK test contained perspectives unknown to the AvaWeb, we believe the low performance (approximately 20% lower IoU) compared to AvaPic and AvaMix is mostly caused by the coarseness of the annotations in combination with low-resolution imagery, which the model struggles to reproduce. However, results also showed that any model trained on avalanches is better than the baseline which has never before seen an avalanche. We believe the coarseness of the annotations in the AvaPic prevents the model from learning all it could from such a large and diverse dataset. Investigating this in more detail is beyond the scope of this paper, but future work should consider experimenting with a larger dataset of finely annotated avalanches covering

various perspectives, avalanche types, avalanche sizes, and snow and illumination conditions.

For their fully automated method, Fox et al. (2023) only evaluated bounding-box overlap, which is less challenging than the pixel overlap we focused on. When comparing our IAS best-model bounding boxes on the first click to the result of Fox et al. (2023), we outperformed their F1 score by a large margin (0.64 vs. 0.97). Consequently, we captured the area that the avalanche covers better from the first prediction onwards.

In our user study, the participants with the best performance were as good as the simulation, but the mean IoU scores of all participants did not exceed the model (Tab. 22.4). We attribute this to the lack of serious training (visible in the variations in the number of clicks and time used) and knowing that estimations of avalanche area exhibit large variabilities (Hafner, 2023) as there is no clear unambiguous definition of an avalanche boundary. Since the differences between the model and the participants were rather small, we consider the way user clicks are simulated during training to be representative of employed real-life click strategies.

Previous work (Hafner, 2023) found variations of up to 43% between experts when mapping avalanches from oblique photographs or from remote sensed imagery. In opposition to Hafner (2023), our mean pairwise IoU scores for the avalanche area mapped (pixels, in our case) were all within 5% of each other and all have an IoU above 0.9 with respect to the ground truth mask (Fig. 22.8). We believe having humans collaborate with the same underlying model homogenizes the avalanche area identified as it guides the participants and constrains the results.

Consequently, IAS not only improves efficiency but enhances the reliability, defined as the consistency of repeated measurements or judgments of the same event relying on the same process (Cronbach, 1947). Even though we had no overlapping avalanches in our UserPic, we still believe our findings also apply in this more challenging scenario.

As opposed to fully automatic avalanche segmentation IAS requires a human annotator. We do not see this as a disadvantage, but rather as complementary since humans are present and will remain present in the future in many settings where avalanches are recorded, either in connection to work or as part of winter leisure activities in the mountains. Compared to the traditional way of mapping avalanches, IAS shows time savings of 90%. We believe that the time saved may be even greater since the avalanches with a time recording were rather small (mean size of 1.75; European avalanche size definition (EAWS, 2023)) compared to the ones in the user study, and all were located in an area well known to the person mapping. In practice, when using the tool to segment new avalanches, the user needs to decide when the predicted and corrected mask are detailed enough. Consequently, the final masks are the most important.

Webcams have limited coverage and cannot record avalanches in a spatially continuous manner like satellite imagery can (Bühler et al., 2019; Eckerstorfer et al., 2019; Hafner et al., 2022), but their temporal resolution is superior and allows for a better monitoring of the avalanche activity over the course of the winter, leading to more complete datasets and allowing for more detailed analyses of, e.g., predisposition factors. However, their inclusion in existing databases, requires the georeferencing of the avalanches, achievable with, e.g., mono-photogrammetry tools like those of Bozzini et al. (2012, 2013), Produit et al. (2016) or Golparvar and Wang (2021). The georeferencing allows for avalanches segmented in an image to be displayed on a map (as is exemplarily shown in Fig. 23.1). Without that, the application is limited to providing an overview of the current activity to an avalanche warning service, while all other downstream applications cannot profit from the data.

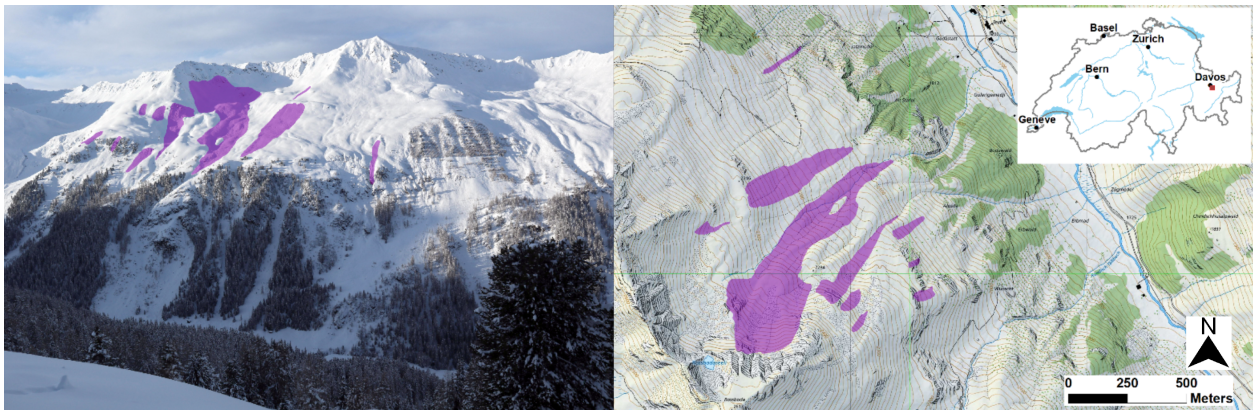


Figure 23.1: Example of avalanches segmented from an image with AvaWeb (left) and the corresponding avalanches displayed on a map after they have been georeferenced with the mono-plotting tool (right, Bozzini et al. (2012)); map source: Federal Office of Topography).

## Conclusions and Outlook

We introduce a novel approach to map avalanches from webcam imagery employing interactive object segmentation. During training, the user's clicks that guide and correct the segmentation were simulated, optimizing the model to quickly identify the features of an avalanche. With IAS, a human user may, in seconds instead of minutes, segment the desired avalanche in collaboration with the model. Compared to satellite imagery, webcam imagery covers only limited areas. However, the abundance of webcams and possibility of acquiring images as frequently as needed without additional cost increase the likelihood of capturing avalanches, even under adverse visibility conditions, offering a very valuable complementary data source for existing avalanche databases. This allows for the documentation of avalanche activity for a whole season compared to for just one extreme event, like in Bühler et al. (2019). Additionally, the release time may be determined with less uncertainty, helping avalanche warning services and research endeavors to better connect the snow and weather conditions to avalanche releases.

In combination, IAS and georeferencing have great potential to improve avalanche mapping: existing mono-photogrammetry tools may be used to import avalanches detected with IAS from webcams. Assuming the camera position and area captured are stable, the georeferencing can be reused for all subsequent images. In the past, this has been done for webcam-based snow cover monitoring (Portenier et al., 2020). In the future, existing approaches could be enhanced and expanded to a pipeline hosting the entire process from IAS to georeferencing and for purposes of importing the detected avalanches into existing databases. Furthermore, we see potential to automatically georeference images from mobile devices with the available information on the location and orientation in combination with the visible skyline and a digital elevation model (DEM). This would allow avalanche observers and the interested backcountry skiers to photograph an observed avalanche; quickly segment it with IAS; and automatically send the georeferenced outlines to existing databases, making them available to, e.g., the avalanche warning service. This would make the outlines and geolocations of avalanches mapped in the field more reliable compared to the traditional mapping approach described in Hafner (2023). The possibility of recording observed avalanches in an easy way could also help to motivate more people in reporting observed avalanches and therefore enlarge current databases with valuable detailed records.

Compared to the currently widely used mapping method (study 2; Hafner, 2023), segmenting an avalanche with IAS shows time savings of over 90%, and the results are more reliable in terms of consistency between mappings from different individuals. For the future we recommend training with a larger dataset with fine annotations and various perspectives, avalanche types, avalanche sizes, and snow and illumination conditions. Our results indicate that this would help the model significantly in making fast and detailed segmentations, as well as in generalizing well to all sorts of unseen perspectives. For fast image annotation or the correction of existing annotations with minimum user input, our current model may be used. Annotations generated with IAS may, in addition, be used to develop and enhance models for automatic avalanche segmentation, saving time while generating outlines that follow the visible avalanche textures, thereby easing the learning and obtaining more accurate and reliable avalanche annotations in the future. Overall, this is a promising approach for continuous and precise avalanche documentation, complementing existing databases and thereby providing a better base for safety-critical decisions and planning in avalanche-prone mountain regions.

# **Part V**

## **Conclusions**



## Contributions of this thesis

This thesis explores the potential of machine learning to automatically map avalanche outlines in remotely sensed imagery. High resolution optical satellites are currently the only source for acquiring comprehensive information on the complete avalanche outlines over a large area (Tab. 5.1). All other currently available methods do not possess the capability to comprehensively cover large areas or to capture the entire extent of avalanches from their release to the deposit area. Webcam imagery, though less comprehensive, has a higher temporal resolution than satellite imagery and is therefore less weather dependent capturing avalanches on selected slopes at regular sub-daily intervals. The information regarding the extent of avalanches obtained from optical satellite imagery or webcams is essential to complement existing avalanche databases used by safety critical applications. For applications such as hazard mapping, planning mitigation measures, or conducting risk analysis, the history of avalanches serves as crucial information when assessing the risk or determining appropriate measures to ensure safety. In avalanche modeling, it is crucial to ensure that the models accurately reflect reality by calibrating them with real avalanches.

This thesis is the first to automate avalanche mapping from optical imagery with deep learning. This work has shown that large scale automatic avalanche mapping from optical SPOT 6/7 satellite imagery is possible, fast and reliable. In reproducibility experiments, the thesis illustrates and quantifies, for the first time, how the perception of avalanches and the delineation of the area covered by one-and-the-same avalanche differs between individuals. In addition, the thesis investigates how the reliability of the human avalanche mapping affects the results of deep learning models and what causes the variations in human estimates. The implications of these investigations for the reliability of avalanche size estimates and mapped avalanche outlines from photographs or remotely sensed imagery are essential for future automation and will influence the perception and management of currently used data. Beyond the field of avalanche research, this largest avalanche mapping reliability investigation to date highlights the importance of understanding reliability of any manually labelled data. With the interactive deep-learning aided mapping of avalanches from webcam imagery, this thesis proposes a second promising way to map avalanches in a fast and reliable way with human guidance, unlocking another data source, besides optical satellite imagery, for the mapping of entire avalanche outlines.

## 25.1 Benefits to practice, science and relevance to society

The methods proposed in this thesis for avalanche mapping provide comprehensive avalanche information over large(r) areas and detailed information on the location, extent, aspect and size of each avalanche (Tab. 5.1). This information is essential for numerical avalanche simulations, hazard mapping, protection/mitigation measure planning and evaluation as well as risk analysis and establishing safety concepts (Tab. 4.1). The proposed methods are not well-suited for avalanche warning services, where rapid notification of avalanche occurrences under all weather conditions is of utmost importance. The data is also of limited use for the development of avalanche models, which connect the weather and snow conditions to avalanche release, because knowing the exact time of release is key (e.g., Kristensen and Larsson, 1994) and our methodology can only be used in conjunction with weather independent avalanche sensors (Tab. 5.1).

It must be mentioned, however, that SAR data like Sentinel-1 is not capable of providing continuous information on avalanche occurrences either: over mid-latitudes, data is currently only available every 12 days. This time-lag is too great for avalanche warning services and for all other applications the data is of little use as it does not provide the relevant information on the complete avalanche outlines (Tab. 4.1 and 5.1). The avalanches mapped from optical satellite imagery can, however, be used to evaluate the avalanche forecast in hindcast, by allowing a comparison of forecasted and documented avalanche activity. In summary, the majority of safety-related applications rely on information about avalanches to make informed decisions, and therefore, benefit from the methods proposed in this thesis. Consequently, this thesis contributes to enhancing safety in mountainous regions for society as a whole.

## 25.2 Limitations

An operational automatic mapping of complete avalanches is currently not possible, as no suitable optical satellite data is available on a regular basis. The spatial resolution of optical satellites like Sentinel-2 (10×10 m), acquiring imagery regularly, is too coarse for a comprehensive mapping, allowing only for the documentation of large glide snow avalanches or slush flows (Hafner et al., 2021; Abermann et al., 2019). Consequently, suitable optical data currently needs to be specifically ordered, acquired and paid for. This is expensive and requires precise knowledge of the specific area of interest, in addition to being reliant on acquiring imagery under cloud-free conditions. The cost, which depends on the covered area, has prevented the testing of our model on data from multiple dates in winter with a large variety of snow conditions and imagery with some clouds. Therefore, further investigation into generalizability is warranted to ensure that the model functions effectively under all possible conditions and times when avalanches may occur.

For the reliability study, the sample for the two mapping experiments was rather small: Due to the homogeneous background and familiarity with the mapping procedure, we judge our results to represent the upper reliability limit. The effect of different countries of origin, self assessed experience or familiarity with the mapping region remains unknown. Consequently, the validity of our findings should be tested with more participants, including various backgrounds. For better comparability and evaluation of all available methods (Tab. 4.1), the reliability of avalanches mapped in other data sources, e.g. from SAR imagery, would be of interest.

The workflow for the IAS from webcam imagery is currently not operational, as it has not been joined with georeferencing to a single model chain. Additionally, the current webcam data is not accessible in the same interface and there is no direct connection to import the avalanches into the designated databases. For operational use, it would also be necessary to include more

finely-mapped avalanches from a wider range of perspectives during training. This would enable the easy segmentation of avalanches from any webcam perspective in the operational setting.

# Avalanche mapping in the future

## 26.1 Avalanches in a changing climate

Matiu et al. (2021) investigated the trends in snow depth in the European Alps from 1971 to 2019, relying on measured data. From May to November, the average snow depth decreased by 8.4%, and the average duration of seasonal snow cover decreased by 5.6%, as assessed across all observed stations. The decrease of the snow cover duration was significant below 2000 m a.s.l., while no consistent change was observed for higher altitudes (Matiu et al., 2021). Marty et al. (2023) described the series of low-snow winters since the 1990s as unprecedented. The observations show less and wetter snow, with increasing variability and an increase in winter melting (Marty et al., 2023). The availability of snow will influence the avalanche activity in the future.

Mayer et al. (2023a) downscaled climate projections (eight climate model chains and three emission scenarios) for two automatic weather stations in Davos and simulated snow stratigraphy until the end of the century. They found a shift of the wet-snow instability to earlier winter months, a decrease in dry snow instability and an increase for wet-snow instabilities at the higher elevation site. Similarly, Ortner et al. (2023) found a general trend towards less area being affected by avalanches as well as reduced avalanche pressure and shorter run-outs when simulating the snowpack of 100 future winters (RCP 8.5 emission scenario considering six different climate model chains). Consequently, the hazard of avalanches will persist at higher elevations and it will lessen at low elevations, while the development at mid-elevations is less sure. In order to capture long-term changes and trends in activity, the comprehensive documentation of avalanches from now into the future is of key relevance.

## 26.2 Deep learning aided avalanche mapping

The volume of data when utilizing comprehensive satellite data or sub-daily webcam imagery to map avalanches is substantial and can only be effectively managed through automation. Deep learning can deal with large datasets due to its ability to automatically learn and extract complex patterns, features and representations directly from the input imagery. In addition, it can scale from small to larger datasets and, unlike humans, is reproducible, yielding identical results when based on the same data and parameters. This makes it a powerful set of methods to effectively and computationally efficiently extract relevant information from large datasets. A future

development for deep learning and avalanche mapping might be the fusion of different data sources and sensors: possible combinations include SAR and optical satellite data, higher and lower resolution optical data as well as a combination of ground-based and spatial/aerial-based sensors. For data with a high temporal resolution, like webcam imagery, recurrent CNNs (e.g., in Daudt et al., 2023) learning from the evolution and occurred changes might be helpful, especially when fully automating the avalanche segmentation. Computer-aided analysis has already been implemented for the Swiss avalanche warning service (e.g., Pérez-Guillén et al., 2022; Hendrick et al., 2023; Mayer et al., 2023b) and has been proposed for data from other sensors (e.g., Heck et al., 2018a; Bianchi et al., 2021; Fox et al., 2023). It can be anticipated that its significance will increase as we continue to deal with ever larger datasets in the future.

### **26.3 Anticipated advancements of sensors and techniques**

Using optical satellite imagery is currently the only option for a comprehensive large-scale mapping of the complete avalanche outlines (for smaller areas, airplane data works as well). Even though suitable data is currently only available when ordered, recorded and paid, the current development for optical satellites is promising. In the future, microsattellites and swarms of satellites will likely provide regular suitable optical imagery at reasonable - lower than nowadays - pricing. The dependence on cloud-free conditions will be mitigated by daily overpasses and will enable regular automatic avalanche mapping over large regions. Regular mappings will improve the understanding of avalanche activity, not only during extreme situations. Currently, the complete outlines of avalanches cannot be derived from weather independent SAR imagery. Radar microsattellites, like ICEYE (ICEEYE, 2024), with higher spatial resolution could make the identification of whole avalanches from SAR imagery possible, though the suitability has yet to be investigated and shown.

In the future, autonomous drones might be able to cover large areas and could be employed to document avalanche periods. The big advantage of drone and aerial imagery is that information on the snow distribution is collected and consequently precise information on the fracture depth and the mass balance/volume distribution can be derived. These drones could be deployed upon request, autonomously documenting avalanche periods. Theoretically, the orthorectification and the automatic avalanche mapping could already be done on-the-fly with an on-board computer, which would make data provision significantly faster.

Webcam imagery allows for a systematic mapping of all avalanche occurrences in the slopes covered by the webcam. Today, most tourist destinations in the Alps maintain webcams, enabling their guests to explore the destination before visiting. Additionally, these webcams serve as a marketing tool, showcasing the beauty of the destination to potential visitors. All webcams capturing avalanche prone slopes could be used in the future to map avalanches with IAS and enlarge the decision base for applications. Even though the spatial coverage of webcams is limited and patchy compared to satellite imagery, they would provide valuable information with superior temporal resolution. Moving from IAS to a reliable, fully automatic avalanche mapping procedure would be desirable for the large-scale use of webcams. Annotating training data with the current IAS interface could help to develop such models. Another future application for IAS would be the combination with automatic georeferencing of avalanche images taken from mobile phones: avalanche observers and interested backcountry recreationists could photograph an observed avalanche with their mobile phone, quickly segment it with IAS and automatically send the georeferenced outlines to existing databases. Future research could explore this, making a pipeline and developing an app for observer guided avalanche mapping with IAS in the field.

The greatest benefit would be cooperative sensors, a combination of currently available sensors

to provide all relevant information for the stakeholders at once (Tab. 4.1 and 5.1). For example, as soon as the weather permits a drone could automatically document the avalanche registered by radar, seismics or infrasound during the snowfall. Likewise, drones could be employed to artificially release avalanches with explosives and consequently map them in detail.

## Final remarks

Capturing regular avalanche activity and especially extreme events as comprehensively as possible continues to be important. This can be achieved by regularly testing new sensors for suitability as well as fully automating information extraction from currently available sensors and data. Furthermore, developing new models and evaluating existing and newly emerging deep learning methods are key for continuously extracting the relevant information from all datasources and passing them on to the relevant stakeholders. In the past, selected avalanche periods have been documented with airplane imagery, for example in Switzerland in the winter of 1950/51 and more systematically in the “avalanche winter” of 1999. Automating the mapping and documenting all avalanches from the available historical data would be extremely important. The available historical data tends to be from extraordinary avalanche periods with larger-than-usual avalanches that are especially important for the downstream applications. Extracting those avalanches would immediately provide a better decision base for the relevant stakeholders, without having to wait for new avalanches to be recorded.

To summarize, enlarging the knowledge about avalanche occurrences with, but not limited to, the proposed methods and automating the information extraction from various sources (Tab. 5.1) is of paramount importance to ensure safety in mountainous regions. Deep learning holds the potential to handle large datasets effectively and produce reliable results, allowing for the identification of patterns induced by climate change and the implementation of appropriate measures. As the population grows and cargo transportation continues, along with the ongoing significance of tourism, having comprehensive and reliable information on avalanche occurrences remains essential for maintaining safety in Switzerland’s mountainous regions.

# **Part VI**

## **Backmatter**





## Appendix to section 6.2

### Short explanation of important terms for CNNs

#### **Backbone**

The backbone refers to the feature extracting network used within the chosen model architecture. It is used to encode the input of the model into a certain feature representations. A typical backbone might be a ResNet (He et al., 2016).

#### **Batch size**

Batch size describes the number of training samples in one iteration (forward-backward pass). Larger batch sizes require more memory as more samples will be propagated through the network. Networks are often trained with mini-batches where the batch size is smaller than the overall number of all samples.

#### **Convergence**

When the loss changes are very small between iterations, a network is converging.

#### **Data augmentation**

Through augmentation like flipping, rotating or adding noise the sample size is increased.

#### **Dilated convolution**

A dilated convolution is a convolution with spacing between the values of the kernel, a kernel of  $3 \times 3$  with a dilation rate of two will cover the same area as a  $5 \times 5$  kernel, but needing nine instead of 25 parameters.

#### **Encoder-decoder**

The encoder takes input data and produces a condensed representation of it, while the decoder takes this representation and generates an feature map with the closest match to the actual input. During training, the model learns to map input to output representations by optimizing parameters to minimize the difference between predicted and actual results. Both encoder and decoder are networks, usually the decoder has the same structure as the encoder but with different direction.

#### **Epoch**

An epoch describes the number of passes through the used training set, one epoch consequently is one forward and one backward pass of all the samples in the training dataset.

#### **Hyper-parameter**

Hyper-parameters like the learning rate or batch size are controlling the learning process and are defined before training.

**Kernel**

A kernel is a small matrix used to extract features from the input data by sliding over it and performing element-wise multiplications and summations to produce feature maps (e.g., in convolutions).

**Learning rate**

The learning rate is a hyper-parameter set before training determining the step size for each iteration when moving towards a minimum of the chosen loss function.

**Loss function**

A loss function is quantifies the difference between the predicted outputs of a model and the target outputs. Through that it guides the model while training and optimizes it.

**Padding**

Padding determines how the border regions of each sample are handled for a convolutional layer. If no padding is applied, part of the border of each kernel (larger than one) will be cropped. away.

**Patch size**

The patch size is the size of the input image processed by the convolutional layers of the network. It determines the area of the input that is considered by each filter during the convolution operation. Larger areas of input/ patch sizes capture more contextual information in contrast to smaller patch sizes which focus on details.

**Receptive field**

The receptive field describes the size of the region in the input that produces the feature (Araujo et al., 2019). A larger receptive field allows the network to capture more context and may improve its ability to generalize to unseen data.

**Strided convolution**

In a standard convolution the kernel will move one pixel at a time (stride = 1). With a stride larger than one the sample image will be downsampled, leading to higher computational efficiency.

**Training, validation and test dataset**

The annotated dataset is usually divided, the largest part is used for training the model. After each training epoch, performance of the model is evaluated on the validation dataset. The test dataset, which the model has not seen during training or validation, is used to evaluate the final performance of the trained model.

## Appendix to part III

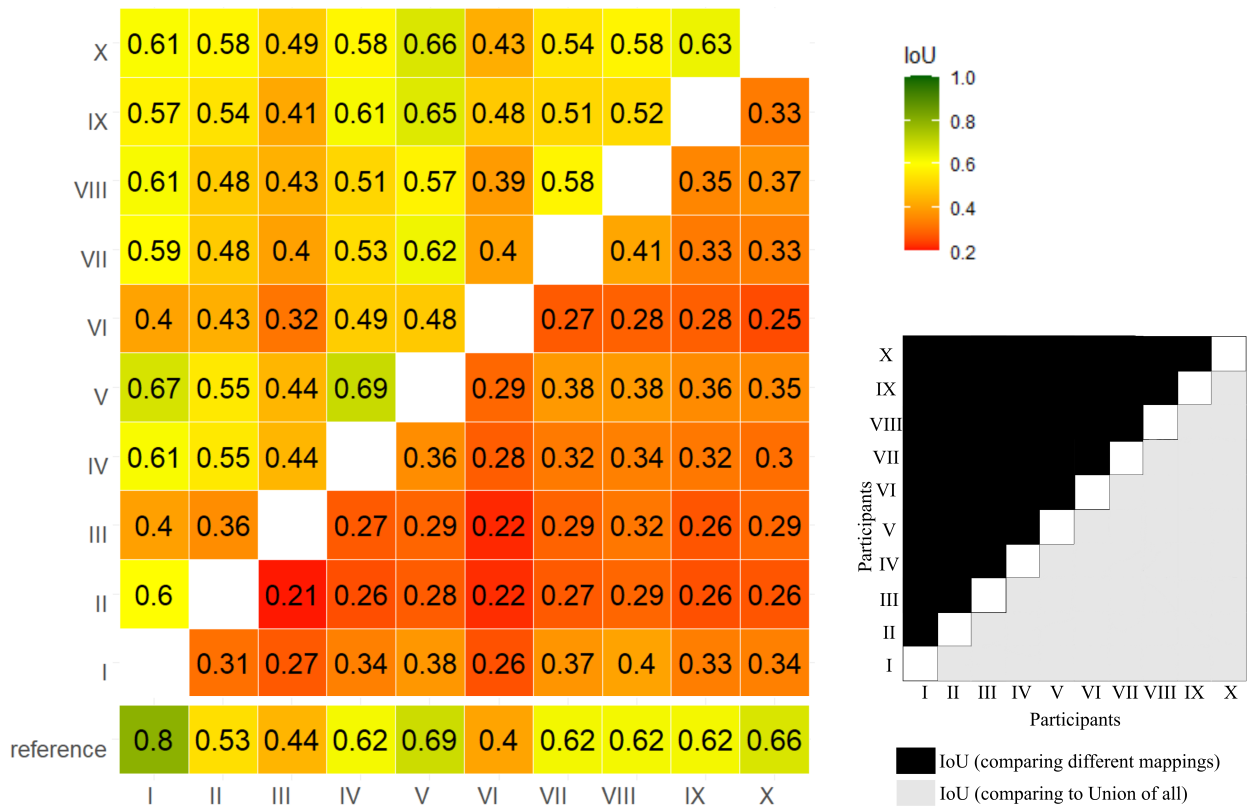


Figure B.1: IoU for all expert pairs for the mapping from oblique photographs. The numbers I to X represent the different expert participants.

## Supplement to part III

Overview on how the survey looked like for the participants. The structure of Question 2 to 10 is identical to Question 1, the images for those questions are displayed in Figure 16.1.

Structure of the survey sent to the forecasters. Question 1 was asked 10 times for each avalanche part of the survey. The survey was created with ArcGIS Survey123.



With this survey we would like to find out how consistently "experts" assess avalanche size. We have prepared eleven questions that will take no more than 10 minutes to answer

Let's go!

**In what role do you deal with avalanches?\***

<input checked="" type="radio"/> Avalanche warner	<input type="radio"/> Mountain guide, tour leader or similar	<input type="radio"/> Observer SLF
<input type="radio"/> Other		

**In which of the following countries do you work for the avalanche warning?\***

If you work as an avalanche warner in a country that is part of the EAWS please use [this survey](#) instead.

<input type="radio"/> Unites States	<input type="radio"/> Canada	<input type="radio"/> other
-------------------------------------	------------------------------	-----------------------------

## Question 1

How big would you estimate this avalanche to be?\*



small avalanche  
(size 1)

medium  
avalanche (size 2)

large avalanche  
(size 3)

very large  
avalanche (size 4)

extremely large  
avalanche (size 5)

do not know

Would you have liked to give an intermediate size?\*

Yes

No

Which one?\*

-Please select-

Size 1 to 2

Size 2 to 3

Size 3 to 4

Size 4 to 5

### Question 11

**How important were the following criteria for you to estimate the avalanche size from the photos?**

(criteria in alphabetic order)

	very important	important	less important	not at all important
<b>Avalanche dimensions*</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Destructive potential*</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Run out*</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Volume*</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Back](#)

[Submit](#)

Powered by ArcGIS Survey123

# List of Publications

The following peer-reviewed publications are associated with this thesis:

- **Hafner, E. D.**, Barton, P., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Automated avalanche mapping from SPOT 6/7 satellite imagery with deep learning: results, evaluation, potential and limitations, *The Cryosphere*, 16, 3517–3530, <https://doi.org/10.5194/tc-16-3517-2022>, 2022.
- Daudt, R. C., Wulf, H., **Hafner, E. D.**, Bühler, Y., Schindler, K., and Wegner, J. D.: Snow depth estimation at country-scale with high spatial and temporal resolution, *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 105–121, <https://doi.org/10.1016/j.isprsjprs.2023.01.017>, 2023.
- Baumer, J., Metzger, N., **Hafner, E. D.**, Daudt, R. C., Wegner, J. D., and Schindler, K.: Automatic Image Compositing and Snow Segmentation for Alpine Snow Cover Monitoring, in: *2023 10th IEEE Swiss Conference on Data Science (SDS)*, pp. 77–84, <https://doi.org/10.1109/SDS57534.2023.00018>, 2023.
- Bührlé, L. J., Marty, M., Eberhard, L. A., Stoffel, A., **Hafner, E. D.**, and Bühler, Y.: Spatially continuous snow depth mapping by aeroplane photogrammetry for annual peak of winter from 2017 to 2021 in open areas, *The Cryosphere*, 17, 3383–3408, <https://doi.org/10.5194/tc-17-3383-2023>, 2023.
- **Hafner, E. D.**, Techel, F., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice, *Natural Hazards and Earth System Sciences*, 23, 2895–2914, <https://doi.org/10.5194/nhess-23-2895-2023>, 2023.
- **Hafner, E. D.**, Kontogianni, T., Daudt, R.C., Oberson, L., Wegner, J. D., Schindler, K., and Bühler, Y.: Interactive Snow Avalanche Segmentation from Webcam Imagery: results, potential and limitations, *The Cryosphere*, 18, 3807–3823, <https://doi.org/10.5194/tc-18-3807-2024>, 2024.
- Dasser, G., Bickel, V. T., Rüetschi, M., Jacquemart, M., Bavay, M., **Hafner, E. D.**, van Herwijnen, A., and Manconi, A.: Monitoring snow wetness evolution from satellite with Sentinel-1 multi-track composites, *EGUsphere [preprint]*, <https://doi.org/10.5194/egusphere-2024-1510>, 2024.

# Bibliography

- Abermann, J., Eckerstorfer, M., Malnes, E., and Hansen, B. U.: A large wet snow avalanche cycle in West Greenland quantified using remote sensing and in situ observations, *Natural Hazards*, 97, 517–534, <https://doi.org/10.1007/s11069-019-03655-8>, 2019.
- American Avalanche Association: *Snow, Weather and Avalanches: Observation Guidelines for Avalanche Programs in the United States*, American Avalanche Association, 4th edn., ISBN 979-8-218-05765-7, <https://www.americanavalancheassociation.org/swag>, 2022.
- Ancey, C.: Snow Avalanches, in: *Geomorphological Fluid Mechanics*, edited by Balmforth, N. J. and Provenzale, A., pp. 319–338, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-45670-4, [https://doi.org/10.1007/3-540-45670-8\\_13](https://doi.org/10.1007/3-540-45670-8_13), 2001.
- Araujo, A., Norris, W., and Sim, J.: Computing Receptive Fields of Convolutional Neural Networks, *Distill*, <https://doi.org/10.23915/distill.00021>, <https://distill.pub/2019/computing-receptive-fields>, 2019.
- Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, *Natural Hazards and Earth System Sciences*, 2, 3–14, <https://doi.org/10.5194/nhess-2-3-2002>, 2002.
- Badoux, A., Andres, N., Techel, F., and Hegg, C.: Natural hazard fatalities in Switzerland from 1946 to 2015, *Natural Hazards and Earth System Sciences*, 16, 2747–2768, <https://doi.org/10.5194/nhess-16-2747-2016>, 2016.
- Baumer, J., Metzger, N., Hafner, E. D., Daudt, R. C., Wegner, J. D., and Schindler, K.: Automatic Image Compositing and Snow Segmentation for Alpine Snow Cover Monitoring, in: *2023 10th IEEE Swiss Conference on Data Science (SDS)*, pp. 77–84, <https://doi.org/10.1109/SDS57534.2023.00018>, 2023.
- Bebi, P., Kulakowski, D., and Rixen, C.: Snow avalanche disturbances in forest ecosystems—State of research and implications for management, *Forest Ecology and Management*, 257, 1883–1892, <https://doi.org/10.1016/j.foreco.2009.01.050>, 2009.
- Bebi, P., Seidl, R., Motta, R., Fuhr, M., Firm, D., Krumm, F., Conedera, M., Ginzler, C., Wohlgemuth, T., and Kulakowski, D.: Changes of forest cover and disturbance regimes in the mountain forests of the Alps, *Forest Ecology and Management*, 388, 43–56, <https://doi.org/10.1016/j.foreco.2016.10.028>, *ecology of Mountain Forest Ecosystems in Europe*, 2017.
- Bedard Jr, A., Greene, G., Intrieri, J., and Rodriguez, R.: On the feasibility and value of detecting and characterizing avalanches remotely by monitoring radiated sub-audible atmospheric sound



- at long distances, *Proceedings of A Multidisciplinary Approach to Snow Engineering*, pp. 267–275, 1988.
- Benenson, R., Popov, S., and Ferrari, V.: Large-Scale Interactive Object Segmentation With Human Annotators, in: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 692–11 701, <https://doi.org/10.1109/CVPR.2019.01197>, 2019.
- BFF and EISLF: Richtlinien zur Berücksichtigung der Lawinengefahr bei Raumwirksamen Tätigkeiten, Bundesamt für Forstwesen (BFF), Eidg. Institut für Schnee- und Lawinenforschung (EISLF), 1984.
- Bianchi, F. M., Grahn, J., Eckerstorfer, M., Malnes, E., and Vickers, H.: Snow Avalanche Segmentation in SAR Images With Fully Convolutional Neural Networks, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 75–82, <https://doi.org/10.1109/JSTARS.2020.3036914>, 2021.
- Birkeland, K. and Greene, E.: Accurately Assessing Avalanche Size: The Ins and Outs of the R- and D- scales., *The Avalanche Review*, 29, 27/32, [https://avalanche.org/wp-content/uploads/2018/08/11\\_TAR\\_BirkelandGreene.pdf](https://avalanche.org/wp-content/uploads/2018/08/11_TAR_BirkelandGreene.pdf), last access: 05.01.2023, 2011.
- Bonnefoy, M., Richard, D., Barral, L., España, J., and Gaucher, R.: The localization map of avalanche phenomena (CLPA): stakes and prospects, in: *12th Congress Interpraevent 2012- Grenoble, France*, pp. 447–457, [https://www.interpraevent.at/palm-cms/upload\\_files/Publikationen/Tagungsbeitraege/2012\\_1\\_447.pdf](https://www.interpraevent.at/palm-cms/upload_files/Publikationen/Tagungsbeitraege/2012_1_447.pdf), last access: 13.02.2024, 2012.
- Boureau, Y.-L., Ponce, J., and Lecun, Y.: A theoretical analysis of feature pooling in visual recognition, *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, p. 111 – 118, <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77956502203&partnerID=40&md5=9c67be5dfa55b4dfcc58e0f4be13d0da>, 2010.
- Bowler, N. E.: Explicitly Accounting for Observation Error in Categorical Verification of Forecasts, *Monthly Weather Review*, 134, 1600 – 1606, <https://doi.org/10.1175/MWR3138.1>, 2006.
- Boykov, Y. and Jolly, M.-P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images, in: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, pp. 105–112 vol.1, <https://doi.org/10.1109/ICCV.2001.937505>, 2001.
- Bozhinskiy, A. N. and Losev, K. S.: The fundamentals of avalanche science, *Mitteilungen des Eidg. Institutes für Schnee- und Lawinenforschung: Vol. 55*. Davos: Eidg. Institut für Schnee- und Lawinenforschung., ISBN 3-905620-71-5, 1998.
- Bozzini, C., Conedera, M., and Krebs, P.: A New Monoplotting Tool to Extract Georeferenced Vector Data and Orthorectified Raster Data from Oblique Non-Metric Photographs, *International Journal of Heritage in the Digital Era*, 1, 499–518, <https://doi.org/10.1260/2047-4970.1.3.499>, 2012.
- Bozzini, C., Conedera, M., and Krebs, P.: A new tool for facilitating the retrieval and recording of the place name cultural heritage, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W2, 115–118, <https://doi.org/10.5194/isprsarchives-XL-5-W2-115-2013>, 2013.

- Brardinoni, F., Scotti, R., Sailer, R., and Mair, V.: Evaluating sources of uncertainty and variability in rock glacier inventories, *Earth Surface Processes and Landforms*, 44, 2450–2466, <https://doi.org/10.1002/esp.4674>, 2019.
- Bründl, M. and Margreth, S.: Integrative Risk Management, in: *Snow and Ice-Related Hazards, Risks and Disasters 2015*, pp. 263–301, W. Haeberli & C. Whiteman (Eds.), <https://doi.org/10.1016/B978-0-12-394849-6.00009-3>, 2015.
- Bründl, M. and Margreth, S.: Chapter 9 - Integrative risk management: The example of snow avalanches, in: *Snow and Ice-Related Hazards, Risks, and Disasters (Second Edition)*, edited by Haeberli, W. and Whiteman, C., *Hazards and Disasters Series*, pp. 259–296, Elsevier, second edition edn., ISBN 978-0-12-817129-5, <https://doi.org/10.1016/B978-0-12-817129-5.00002-0>, 2021.
- Bründl, M., Hafner, E. D., Bebi, P., Bühler, Y., Margreth, S., Marty, C., Schaer, M., Stoffel, L., Techel, F., Winkler, K., Zweifel, B., and Schweizer, J.: Ereignisanalyse Lawinsituation im Januar 2018, vol. 76 of *WSL Berichte*, Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL, Birmensdorf, <https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A19842>, last access: 16.02.2024, 2019.
- Bründl, M.: Dealing with Natural Hazard Risks in Switzerland - The Influence of Hazard Mapping on Risk-based Decision Making, in: *Advances in Global Change Research: Dating Torrential Processes on Fans and Cones. Methods and Their Application for Hazard and Risk Assessment*, vol. 47 of *Advances in Global Change Research*, pp. 355–365, Schneuwly-Bollschweiler, M. and Stoffel, M. and Rudolf-Miklau, F. (eds.), [https://doi.org/10.1007/978-94-007-4336-6\\_24](https://doi.org/10.1007/978-94-007-4336-6_24), 2013.
- Bühler, Y., Hüni, A., Christen, M., Meister, R., and Kellenberger, T.: Automated detection and mapping of avalanche deposits using airborne optical remote sensing data, *Cold Regions Science and Technology*, 57, 99–106, <https://doi.org/10.1016/j.coldregions.2009.02.007>, 2009.
- Bühler, Y., Adams, M. S., Bösch, R., and Stoffel, A.: Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations, *The Cryosphere*, 10, 1075–1088, <https://doi.org/10.5194/tc-10-1075-2016>, 2016.
- Bühler, Y., Adams, M. S., Stoffel, A., and Boesch, R.: Photogrammetric reconstruction of homogenous snow surfaces in alpine terrain applying near-infrared UAS imagery, *International Journal of Remote Sensing*, 38, 3135–3158, <https://doi.org/10.1080/01431161.2016.1275060>, 2017.
- Bühler, Y., Hafner, E. D., Zweifel, B., Zesiger, M., and Heisig, H.: Where are the avalanches? Rapid SPOT6 satellite data acquisition to map an extreme avalanche period over the Swiss Alps, *The Cryosphere*, 13, 3225–3238, <https://doi.org/10.5194/tc-13-3225-2019>, 2019.
- Bühler, Y., Bebi, P., Christen, M., Margreth, S., Stoffel, L., Stoffel, A., Marty, C., Schmucki, G., Caviezel, A., Kühne, R., Wohlwend, S., and Bartelt, P.: Automated avalanche hazard indication mapping on a statewide scale, *Natural Hazards and Earth System Sciences*, 22, 1825–1843, <https://doi.org/10.5194/nhess-22-1825-2022>, 2022.
- Bührle, L. J., Marty, M., Eberhard, L. A., Stoffel, A., Hafner, E. D., and Bühler, Y.: Spatially continuous snow depth mapping by aeroplane photogrammetry for annual peak of winter from

- 2017 to 2021 in open areas, *The Cryosphere*, 17, 3383–3408, <https://doi.org/10.5194/tc-17-3383-2023>, 2023.
- Butler, D.: Spatial and temporal aspects of the snow avalanche hazard, Glacier National Park, Montana, U.S.A., in: *International Snow Science Workshop Proceedings 1986, Lake Tahoe, California, USA*, pp. 223–230, <https://arc.lib.montana.edu/snow-science/objects/issw-1986-223-230.pdf>, last access: 13.02.2024, 1986.
- CAA: Avalanche Size, <https://www.avalanche-center.org/Education/glossary/avalanche-size.php>, last access: 04.01.2023, 2023.
- Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., and Li, Z.: A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach, *Remote Sensing of Environment*, 210, 35–47, <https://doi.org/10.1016/j.rse.2018.02.045>, 2018.
- Caiserman, A., Sidle, R. C., and Gurung, D. R.: Snow Avalanche Frequency Estimation (SAFE): 32 years of monitoring remote avalanche depositional zones in high mountains of Afghanistan, *The Cryosphere*, 16, 3295–3312, <https://doi.org/10.5194/tc-16-3295-2022>, 2022.
- Canadian Avalanche Association: Observation guidelines and recording standards for weather, snowpack and avalanches, Canadian Avalanche Association, ISBN 0-9685856-3-9, [https://cdn.ymaws.com/www.avalancheassociation.ca/resource/resmgr/standards\\_docs/OGRS2016web.pdf](https://cdn.ymaws.com/www.avalancheassociation.ca/resource/resmgr/standards_docs/OGRS2016web.pdf), 2016.
- Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, *CoRR*, abs/1802.02611, <https://doi.org/10.48550/arXiv.1802.02611>, 2018.
- Chen, L., Strauch, M., and Merhof, D.: Instance Segmentation of Biomedical Images with an Object-aware Embedding Learned with Local Constraints, <https://doi.org/10.48550/arXiv.2004.09821>, 2020.
- Christen, M., Kowalski, J., and Bartelt, P.: RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain, *Cold Regions Science and Technology*, 63, 1–14, <https://doi.org/10.1016/j.coldregions.2010.04.005>, 2010.
- Coaz, J. F., ed.: *Die Lauinen der Schweizeralpen*, Eidgenössisches Handels- und Landwirtschaftsdepartment, Bern, Switzerland, <https://doi.org/10.3931/e-rara-21209>, 1881.
- Conedera, M., Colombaroli, D., Tinner, W., Krebs, P., and Whitlock, C.: Insights about past forest dynamics as a tool for present and future forest management in Switzerland, *Forest Ecology and Management*, 388, 100–112, <https://doi.org/10.1016/j.foreco.2016.10.027>, ecology of Mountain Forest Ecosystems in Europe, 2017.
- Conedera, M., Bozzini, C., Ryter, U., Bertschinger, T., and Krebs, P.: Using the Monoplotting Technique for Documenting and Analyzing Natural Hazard Events, in: *Natural Hazards - Risk Assessment and Vulnerability Reduction*, IntechOpen, <https://doi.org/10.5772/intechopen.77321>, 2018.
- Cronbach, L. J.: Test “reliability”: Its meaning and determination, *Psychometrika*, 12, 1–16, <https://doi.org/10.1007/bf02289289>, 1947.

- Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y.: Deformable Convolutional Networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 764–773, <https://doi.org/10.1109/ICCV.2017.89>, 2017.
- Daudt, R. C., Wulf, H., Hafner, E. D., Bühler, Y., Schindler, K., and Wegner, J. D.: Snow depth estimation at country-scale with high spatial and temporal resolution, *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 105–121, <https://doi.org/10.1016/j.isprsjprs.2023.01.017>, 2023.
- de Bouchard d’Aubeterre, G., Favillier, A., Mainieri, R., Lopez Saez, J., Eckert, N., Saulnier, M., Peiry, J.-L., Stoffel, M., and Corona, C.: Tree-ring reconstruction of snow avalanche activity: Does avalanche path selection matter?, *Science of The Total Environment*, 684, 496–508, <https://doi.org/10.1016/j.scitotenv.2019.05.194>, 2019.
- De Brabandere, B., Neven, D., and Van Gool, L.: Semantic Instance Segmentation for Autonomous Driving, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 478–480, <https://doi.org/10.1109/CVPRW.2017.66>, 2017.
- EAWS: Standards: Avalanche Size, <https://www.avalanches.org/standards/avalanche-size/>, last access: 08.05.2023, 2023.
- EAWS: Standards: Avalanche Danger Scale, <https://www.avalanches.org/standards/avalanche-danger-scale/>, last access: 2024-02-16, 2024a.
- EAWS: Glossary, <https://www.avalanches.org/glossary/>, last access: 2024-03-11, 2024b.
- Eckerstorfer, M. and Malnes, E.: Manual detection of snow avalanche debris using high-resolution Radarsat-2 SAR images, *Cold Regions Science and Technology*, 120, 205–218, <https://doi.org/10.1016/j.coldregions.2015.08.016>, 2015.
- Eckerstorfer, M., Bühler, Y., Frauenfelder, R., and Malnes, E.: Remote sensing of snow avalanches: Recent advances, potential, and limitations, *Cold Regions Science and Technology*, 121, 126–140, <https://doi.org/10.1016/j.coldregions.2015.11.001>, 2016.
- Eckerstorfer, M., Malnes, E., and Müller, K.: A complete snow avalanche activity record from a Norwegian forecasting region using Sentinel-1 satellite-radar data, *Cold Regions Science and Technology*, 144, 39–51, <https://doi.org/10.1016/j.coldregions.2017.08.004>, 2017.
- Eckerstorfer, M., Vickers, H., Malnes, E., and Grahn, J.: Near-Real Time Automatic Snow Avalanche Activity Monitoring System Using Sentinel-1 SAR Data in Norway, *Remote Sensing*, 11, <https://doi.org/10.3390/rs11232863>, 2019.
- Eckerstorfer, M., Oterhals, H. D., Müller, K., Malnes, E., Grahn, J., Langeland, S., and Velsand, P.: Performance of manual and automatic detection of dry snow avalanches in Sentinel-1 SAR images, *Cold Regions Science and Technology*, 198, 103–114, <https://doi.org/10.1016/j.coldregions.2022.103549>, 2022.
- Egger, T., Kadelbach, T., Koch, A., and SAB: Das Schweizer Berggebiet 2020: Fakten und Zahlen, vol. 245, Schweizerische Arbeitsgemeinschaft für die Berggebiete (SAB), [https://www.sab.ch/wp-content/uploads/2022/11/Nr246-BIZ2020\\_08.07.2020.pdf](https://www.sab.ch/wp-content/uploads/2022/11/Nr246-BIZ2020_08.07.2020.pdf), last access: 19.03.2024, 2020.

- EISLF: Der Lawinenwinter 1999: Ereignisanalyse, WSL Institute for Snow and Avalanche Research SLF, ISBN 3-905620-80-4, 2000.
- ESA: DEFROST, <https://business.esa.int/projects/defrost>, last access: 08.02.2024, 2020.
- ESA: SPOT 6 Instruments, <https://earth.esa.int/eogateway/missions/spot-6/description>, last access: 13.03.2024, 2024a.
- ESA: SentiWiki: Copernicus Programme, <https://sentiwiki.copernicus.eu/web/copernicus-programme>, last access: 13.03.2024, 2024b.
- Federal Statistical Office: Regional indicators on the Tourism Satellite Account, <https://www.bfs.admin.ch/bfs/en/home/statistics/tourism/monetary-aspects/regional-indicators.html>, last access: 19.03.2024, 2024.
- Fees, A., van Herwijnen, A., Altenbach, M., Lombardo, M., and Schweizer, J.: Glide-snow avalanche characteristics at different timescales extracted from time-lapse photography, *Annals of Glaciology*, p. 1–12, <https://doi.org/10.1017/aog.2023.37>, 2023.
- Feick, S., Mitterer, C., Dreier, L., Harvey, S., and Schweizer, J.: Automated detection and monitoring of glide-snow events using satellite-based optical remote sensing and terrestrial photography, in: *Proceedings of the 2012 International Snow Science Workshop*, Anchorage AK, USA, pp. 603–609, <https://arc.lib.montana.edu/snow-science/objects/issw-2012-603-609.pdf>, last access: 21.02.2024, 2012.
- Fiorino, C., Reni, M., Bolognesi, A., and Calandrino, R.: Intra- and inter-observer variability in contouring prostate and seminal vesicles: Implications for conformal treatment planning, *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*, 47, 285–92, [https://doi.org/10.1016/S0167-8140\(98\)00021-8](https://doi.org/10.1016/S0167-8140(98)00021-8), 1998.
- Fox, J., Siebenbrunner, A., Reitingner, S., Peer, D., and Rodríguez-Sánchez, A.: Deep Learning for Real-Time Avalanche Detection in Webcam Images, *International Snow Science Workshop ISSW*, Bend, 8-13 October 2023, <https://arc.lib.montana.edu/snow-science/item.php?id=3087>, last access: 25.03.2024, 2023.
- Frauenfelder, R., Solberg, R., Larsen, S. Ø., Salberg, A.-B., and Bjordal, H.: Remote-sensing derived avalanche inventory data, in: *Proceedings of the 2012 International Snow Science Workshop*, Anchorage AK, USA, pp. 398–403, <https://arc.lib.montana.edu/snow-science/objects/issw-2012-398-403.pdf>, last access: 21.02.2024, 2012.
- Frauenfelder, R., Lato, M. J., and Biskupič, M.: Using eCognition to automatically detect and map avalanche deposits from the spring 2009 avalanche cycle in the Tatra mts., Slovakia, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W3, 791–795, <https://doi.org/10.5194/isprsarchives-XL-7-W3-791-2015>, 2015.
- Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., and Reichenbach, P.: Comparing landslide inventory maps, *Geomorphology*, 94, 268–289, <https://doi.org/10.1016/j.geomorph.2006.09.023>, 2008.

- Giacona, F., Eckert, N., and Martin, B.: A 240-year history of avalanche risk in the Vosges Mountains based on non-conventional (re)sources, *Natural Hazards and Earth System Sciences*, 17, 887–904, <https://doi.org/10.5194/nhess-17-887-2017>, 2017.
- Goffin, R. D. and Olson, J. M.: Is It All Relative?: Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others, *Perspectives on Psychological Science*, 6, 48–60, <https://doi.org/10.1177/1745691610393521>, 2011.
- Golparvar, B. and Wang, R.-Q.: AI-supported Framework of Semi-Automatic Monoplotting for Monocular Oblique Visual Data Analysis, <https://doi.org/10.48550/arXiv.2111.14021>, 2021.
- Goodfellow, I., Bengio, Y., and Courville, A.: *Deep Learning*, MIT Press, <http://www.deeplearningbook.org>, last access: 13.03.2024, 2016.
- Gruber, U. and Margreth, S.: Winter 1999: a valuable test of the avalanche-hazard mapping procedure in Switzerland, *Annals of Glaciology*, 32, 328–332, <https://doi.org/10.3189/172756401781819238>, 2001.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., and Chen, T.: Recent advances in convolutional neural networks, *Pattern Recognition*, 77, 354–377, <https://doi.org/10.1016/j.patcog.2017.10.013>, 2018.
- Gulshan, V., Rother, C., Criminisi, A., Blake, A., and Zisserman, A.: Geodesic star convexity for interactive image segmentation, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3129–3136, <https://doi.org/10.1109/CVPR.2010.5540073>, 2010.
- Gupta, A., Dollar, P., and Girshick, R.: LVIS: A Dataset for Large Vocabulary Instance Segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), <https://doi.org/10.48550/arXiv.1908.03195>, 2019.
- Hafner, E. D.: Data reliability study: avalanche size estimates and outlines [data set], <https://doi.org/10.16904/envidat.423>, 2023.
- Hafner, E. D. and Bühler, Y.: SPOT6 Avalanche outlines 24 January 2018 [data set], <https://doi.org/10.16904/envidat.77>, 2019.
- Hafner, E. D. and Bühler, Y.: SPOT6 Avalanche outlines 16 January 2019 [data set], <https://doi.org/10.16904/envidat.235>, 2021.
- Hafner, E. D., Techel, F., Leinss, S., and Bühler, Y.: Mapping avalanches with satellites – evaluation of performance and completeness, *The Cryosphere*, 15, 983–1004, <https://doi.org/10.5194/tc-15-983-2021>, 2021.
- Hafner, E. D., Barton, P., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Automated avalanche mapping from SPOT 6/7 satellite imagery with deep learning: results, evaluation, potential and limitations, *The Cryosphere*, 16, 3517–3530, <https://doi.org/10.5194/tc-16-3517-2022>, 2022.
- Hafner, E. D., Oberson, L., Kontogianni, T., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Using interactive object segmentation to derive avalanche outlines from webcam imagery, in: EGU General Assembly 2023, Vienna, Austria, <https://doi.org/10.5194/egusphere-egu23-10867>, eGU23-10867, 2023a.

- Hafner, E. D., Techel, F., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice, *Natural Hazards and Earth System Sciences*, 23, 2895–2914, <https://doi.org/10.5194/nhess-23-2895-2023>, 2023b.
- Hafner, E. D., Kontogianni, T., Daudt, R. C., Oberson, L., Wegner, J. D., Schindler, K., and Bühler, Y.: Interactive Snow Avalanche Segmentation from Webcam Imagery: results, potential and limitations, *The Cryosphere*, 18, 3807–3823, <https://doi.org/10.5194/tc-18-3807-2024>, 2024.
- Hagafors, R. and Brehmer, B.: Does having to justify one's judgments change the nature of the judgment process?, *Organizational Behavior and Human Performance*, 31, 223–232, [https://doi.org/10.1016/0030-5073\(83\)90122-8](https://doi.org/10.1016/0030-5073(83)90122-8), 1983.
- Hamar, J. B., Salberg, A.-B., and Ardelean, F.: Automatic detection and mapping of avalanches in SAR images, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 689–692, <https://doi.org/10.1109/IGARSS.2016.7729173>, 2016.
- Harrison, J.: Seismic signals from avalanches, in: *Avalanche release and snow characteristics*, edited by Armstrong, R. L. and Ives, J. D., no. 19 in *Occasional Paper*, pp. 145–150, Institute of Arctic and Alpine Research, University of Colorado, <https://snowstudies.org/wp-content/uploads/2020/03/OP19-AVALANCHE-RELEASE-AND-SNOW-CHARACTERISTICS-Reduced.pdf>, last access: 19.02.2024, 1976.
- Harvey, S., Winkler, K., Techel, F., and Marty, C.: Schnee und Lawinen in den Schweizer Alpen. Hydrologisches Jahr 2006/07, WSL-Institut für Schnee- und Lawinenforschung SLF, Davos, <https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A16485>, last access: 20.3.2024, 2013.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>, 2016.
- He, K., Gkioxari, G., Dollár, P., and Girshick, R.: Mask R-CNN, <https://doi.org/10.48550/arXiv.1703.06870>, 2018.
- Heck, M., Hammer, C., Hobiger, M., van Herwijnen, A., Schweizer, J., and Fäh, D.: Automatic classification of continuous seismic data for avalanche monitoring purposes, in: *International Snow Science Workshop Proceedings 2018*, Innsbruck, Austria, pp. 631–635, [https://arc.lib.montana.edu/snow-science/objects/ISSW2018\\_007.10.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW2018_007.10.pdf), last access: 19.02.2024, 2018a.
- Heck, M., Hobiger, M., van Herwijnen, A., Schweizer, J., and Fäh, D.: Localization of seismic events produced by avalanches using multiple signal classification, *Geophysical Journal International*, 216, 201–217, <https://doi.org/10.1093/gji/ggy394>, 2018b.
- Helbig, N., van Herwijnen, A., and Jonas, T.: Forecasting wet-snow avalanche probability in mountainous terrain, *Cold Regions Science and Technology*, 120, 219–226, <https://doi.org/10.1016/j.coldregions.2015.07.001>, 2015.
- Hendrick, M., Techel, F., Volpi, M., Olevski, T., Pérez-Guillén, C., Herwijnen, A. v., and Schweizer, J.: Automated prediction of wet-snow avalanche activity in the Swiss Alps, *Journal of Glaciology*, 69, 1365–1378, <https://doi.org/10.1017/jog.2023.24>, 2023.

- Hendrikx, J., Owens, I., Carran, W., and Carran, A.: Avalanche activity in an extreme maritime climate: The application of classification trees for forecasting, *Cold Reg. Sci. Technol.*, 43, 104–116, 2005.
- Hendrikx, J., Peitzsch, E. H., and Fagre, D. B.: Time-lapse photography as an approach to understanding glide avalanche activity, in: *Proceedings of the 2012 International Snow Science Workshop*, Anchorage, Alaska, USA, pp. 872–877, <https://arc.lib.montana.edu/snow-science/objects/issw-2012-872-877.pdf>, last access: 21.02.2024, 2012.
- Hoffer, E., Ben-Nun, T., Hubara, I., Giladi, N., Hoefler, T., and Soudry, D.: Augment your batch: better training with larger batches, <https://doi.org/10.48550/arXiv.1901.09335>, 2019.
- Holub, M. and Hübl, J.: Local protection against mountain hazards - state of the art and future needs, *Natural Hazards and Earth System Sciences*, 8, 81–99, <https://doi.org/10.5194/nhess-8-81-2008>, 2008.
- ICEEYE: ICEEYE, <https://www.iceeye.com/satellites>, last access: 16.03.2024, 2024.
- Jamieson, B., Haegeli, P., and Schweizer, J.: Field observations for estimating the local avalanche danger in the Columbia Mountains of Canada, *Cold Regions Science and Technology*, 58, 84–91, <https://doi.org/10.1016/j.coldregions.2009.03.005>, 2009.
- Jamieson, B., Beglinger, R., and Wilson, D.: Case study of a large snow avalanche in the Selkirk Mountains and reflections on the Canadian size classification, *Geohazards 6 - 6th Canadian GeoHazards Conference*, Kingston, ON, Canada, <https://cgs.ca/docs/geohazards/kingston2014/Geo2014/pdfs/geoHaz6Paper113.pdf>, last access: 25.03.2024, 2014.
- Kapper, K. L., Goelles, T., Muckenhuber, S., Trügler, A., Abermann, J., Schlager, B., Gaisberger, C., Eckerstorfer, M., Grahn, J., Malnes, E., Prokop, A., and Schöner, W.: Automated snow avalanche monitoring for Austria: State of the art and roadmap for future work, *Frontiers in Remote Sensing*, 4, <https://doi.org/10.3389/frsen.2023.1156519>, 2023.
- Karas, A., Karbou, F., Giffard-Roisin, S., Durand, P., and Eckert, N.: Automatic Color Detection-Based Method Applied to Sentinel-1 SAR Images for Snow Avalanche Debris Monitoring, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17, <https://doi.org/10.1109/TGRS.2021.3131853>, 2022.
- Karbou, F., Coléou, C., Lefort, M., Deschatres, M., Eckert, N., Martin, R., Charvet, G., and Dufour, A.: Monitoring avalanche debris in the French mountains using SAR observations from Sentinel-1 satellites, *International Snow Science Workshop ISSW*, Innsbruck, pp. 344–347, <https://arc.lib.montana.edu/snow-science/item/2548>, last access: 25.03.2024, 2018.
- Keskinen, Z., Hendrikx, J., Eckerstorfer, M., and Birkeland, K.: Satellite detection of snow avalanches using Sentinel-1 in a transitional snow climate, *Cold Regions Science and Technology*, 199, 103 558, <https://doi.org/https://doi.org/10.1016/j.coldregions.2022.103558>, 2022.
- Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, <https://doi.org/10.48550/arXiv.1412.6980>, 2017.
- Kingma, D. P. and Welling, M.: An Introduction to Variational Autoencoders, *Foundations and Trends in Machine Learning*, 12, <https://doi.org/10.48550/arXiv.1906.02691>, 2019.



- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R.: Segment Anything, <https://doi.org/10.48550/arXiv.2304.02643>, 2023.
- Kogelnig, A., Suriñach, E., Vilajosana, I., Hübl, J., Sovilla, B., Hiller, M., and Dufour, F.: On the complementariness of infrasound and seismic sensors for monitoring snow avalanches, *Natural Hazards and Earth System Sciences*, 11, 2355–2370, <https://doi.org/10.5194/nhess-11-2355-2011>, 2011.
- Köhler, A., McElwaine, J. N., and Sovilla, B.: GEODAR Data and the Flow Regimes of Snow Avalanches, *Journal of Geophysical Research: Earth Surface*, 123, 1272–1294, <https://doi.org/10.1002/2017jf004375>, 2018.
- Kontogianni, T., Gygli, M., Uijlings, J., and Ferrari, V.: Continuous Adaptation for Interactive Object Segmentation by Learning from Corrections, <https://doi.org/10.48550/arXiv.1911.12709>, 2020.
- Korzeniowska, K., Bühler, Y., Marty, M., and Korup, O.: Regional snow-avalanche detection using object-based image analysis of near-infrared aerial imagery, *Natural Hazards and Earth System Sciences*, 17, 1823–1836, <https://doi.org/10.5194/nhess-17-1823-2017>, 2017.
- Kosberg, S., Müller, K., Landrø, M., Ekker, R., and Engeset, R.: Key to success for the Norwegian Avalanche Center: Merging of theoretical and practical knowhow, *International Snow Science Workshop ISSW*, [https://arc.lib.montana.edu/snow-science/objects/ISSW13\\_paper\\_P1-21.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW13_paper_P1-21.pdf), last access: 07.06.2024, 2013.
- Koschuch, R.: 8 Years Experience in Avalanche Detection by Using a Pulse Doppler Radar, in: *International Snow Science Workshop Proceedings 2018, Innsbruck, Austria*, pp. 636–639, [https://arc.lib.montana.edu/snow-science/objects/ISSW2018\\_007.12.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW2018_007.12.pdf), last access: 12.02.2024, 2018.
- Kristensen, K. and Larsson, C.: An Avalanche Forecasting Program Based on a Modified Nearest Neighbour Method, in: *Proceedings of the 1994 International Snow Science Workshop, Snowbird, Utah, USA*, pp. 22–30, <https://arc.lib.montana.edu/snow-science/objects/issw-1994-022-030.pdf>, last access: 12.02.2024, 1994.
- Kull, C. A.: Historical landscape repeat photography as a tool for land use change research, *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*, 59, 253–268, <https://doi.org/10.1080/00291950500375443>, 2005.
- Lampert, T., Stumpf, A., and Gancarski, P.: An empirical study into annotator agreement, ground truth estimation, and algorithm evaluation, *IEEE Transactions on Image Processing*, 25, 2557–2572, 2016.
- Latenser, M. and Pfister, C.: Rapid mass movement as a source of climatic evidence for the Holocene: Avalanches in Switzerland 1500-1990, *Journal of Quaternary Science*, 19, 241–266, [https://www.hist.unibe.ch/unibe/portal/fak\\_historisch/dga/hist/content/e11168/e52524/e69145/e186327/e188664/52\\_Latenser\\_pfister\\_Avalanches-97\\_ger.pdf](https://www.hist.unibe.ch/unibe/portal/fak_historisch/dga/hist/content/e11168/e52524/e69145/e186327/e188664/52_Latenser_pfister_Avalanches-97_ger.pdf), last access: 12.03.2024, 1997.
- Lato, M. J., Frauenfelder, R., and Bühler, Y.: Automated detection of snow avalanche deposits: Segmentation and classification of optical remote sensing imagery, *Natural Hazards and Earth System Sciences*, 12, 2893–2906, <https://doi.org/10.5194/nhess-12-2893-2012>, 2012.

- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86, 2278–2324, <https://doi.org/10.1109/5.726791>, 1998.
- Leinss, S., Wicki, R., Holenstein, S., Baffelli, S., and Bühler, Y.: Snow avalanche detection and mapping in multitemporal and multiorbital radar images from TerraSAR-X and Sentinel-1, *Natural Hazards and Earth System Sciences*, 20, 1783–1803, <https://doi.org/10.5194/nhess-20-1783-2020>, 2020.
- Levandowsky, M. and Winter, D.: Distance between sets, *Nature*, 234, 34–35, <https://doi.org/10.1038/234034a0>, 1971.
- Li, Z., Chen, Q., and Koltun, V.: Interactive Image Segmentation with Latent Diversity, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 577–585, <https://doi.org/10.1109/CVPR.2018.00067>, 2018.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P.: Microsoft COCO: Common Objects in Context, <https://doi.org/10.48550/arXiv.1405.0312>, 2015.
- Lin, Z., Duan, Z.-P., Zhang, Z., Guo, C.-L., and Cheng, M.-M.: FocusCut: Diving into a Focus View in Interactive Segmentation, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2627–2636, <https://doi.org/10.1109/CVPR52688.2022.00266>, 2022.
- Liu, Y., Li, H., Hu, C., Luo, S., Luo, Y., and Chen, C. W.: Learning to Aggregate Multi-Scale Context for Instance Segmentation in Remote Sensing Images, 2022.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A.: Deep learning in remote sensing applications: A meta-analysis and review, *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177, <https://doi.org/10.1016/j.isprsjprs.2019.04.015>, 2019.
- Mahadevan, S., Voigtlaender, P., and Leibe, B.: Iteratively Trained Interactive Segmentation, in: 2018 Conference on Computer Vision and Pattern Recognition (CVPR), <https://doi.org/10.48550/arXiv.1805.04398>, 2018.
- Marchetti, E., van Herwijnen, A., Christen, M., Silengo, M. C., and Barfucci, G.: Seismo-acoustic energy partitioning of a powder snow avalanche, *Earth Surface Dynamics*, 8, 399–411, <https://doi.org/10.5194/esurf-8-399-2020>, 2020.
- Margreth, S.: Rezoning after Installing Avalanche Mitigation Measures: Case Study of the Vallascia Avalanche in Airolo, Switzerland, in: *International Snow Science Workshop Proceedings 2018*, Innsbruck, Austria, pp. 111–115, [https://arc.lib.montana.edu/snow-science/objects/ISSW2018\\_002.2.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW2018_002.2.pdf), last access: 13.02.2024, 2018.
- Margreth, S. and Romang, H.: Effectiveness of mitigation measures against natural hazards, *Cold Regions Science and Technology*, 64, 199–207, <https://doi.org/10.1016/j.coldregions.2010.04.013>, *international Snow Science Workshop 2009 Davos*, 2010.
- Marty, C., Rohrer, M. B., Huss, M., and Stähli, M.: Multi-decadal observations in the Alps reveal less and wetter snow, with increasing variability, *Frontiers in Earth Science*, 11, <https://doi.org/10.3389/feart.2023.1165861>, 2023.

- Mather, A. and Fairbairn, J.: From Floods to Reforestation: The Forest Transition in Switzerland, *Environment and History*, 6, 399–421, <https://doi.org/10.3197/096734000129342352>, 2000.
- Mather, A., Fairbairn, J., and Needle, C.: The course and drivers of the forest transition: The case of France, *Journal of Rural Studies*, 15, 65–90, [https://doi.org/10.1016/S0743-0167\(98\)00023-0](https://doi.org/10.1016/S0743-0167(98)00023-0), 1999.
- Mathieu, J., Backhaus, N., Hürlimann, K., and Bürgi, M.: *Geschichte der Landschaft in der Schweiz*, Orell Füssli Verlag, 1 edn., ISBN 978-3-280-05601-1, 2016.
- Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., De Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., Notarnicola, C., Petitta, M., Scherrer, S. C., Strasser, U., Winkler, M., Zebisch, M., Cicogna, A., Cremonini, R., Debernardi, A., Faletto, M., Gaddo, M., Giovannini, L., Mercalli, L., Soubeyroux, J.-M., Sušnik, A., Trenti, A., Urbani, S., and Weilguni, V.: Observed snow depth trends in the European Alps: 1971 to 2019, *The Cryosphere*, 15, 1343–1382, <https://doi.org/10.5194/tc-15-1343-2021>, 2021.
- Mayer, S., van Herwijnen, A., Ulivieri, G., and Schweizer, J.: Evaluating the performance of an operational infrasound avalanche detection system at three locations in the Swiss Alps during two winter seasons, *Cold Regions Science and Technology*, 173, 102–962, <https://doi.org/10.1016/j.coldregions.2019.102962>, 2020.
- Mayer, S., Hendrick, M., Michel, A., Richter, B., Schweizer, J., and van Herwijnen, A.: Projected impact of climate warming on avalanche activity at two sites in the Swiss Alps, in: *International Snow Science Workshop Grenoble – Bend, Oregon - 2023*, pp. 572–576, [https://arc.lib.montana.edu/snow-science/objects/ISSW2023\\_06.05.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW2023_06.05.pdf), last access: 13.02.2024, 2023a.
- Mayer, S., Techel, F., Schweizer, J., and van Herwijnen, A.: Prediction of natural dry-snow avalanche activity using physics-based snowpack simulations, *Natural Hazards and Earth System Sciences*, 23, 3445–3465, <https://doi.org/10.5194/nhess-23-3445-2023>, 2023b.
- McClung, D. and Schaerer, P.: *The Avalanche Handbook*, Mountaineers Books, Seattle, WA, 3 edn., ISBN 0-89886-809-2, 2006.
- McClung, D. M. and Schaerer, P. A.: Snow avalanche size classification, *Avalanche Workshop*, pp. 12–27, <https://arc.lib.montana.edu/snow-science/objects/issw-1980-012-030.pdf>, last access: 11.11.2022, 1980.
- McClung, D. M. and Schaerer, P. A.: Characteristics of Flowing Snow and Avalanche Impact Pressures, *Annals of Glaciology*, 6, 9–14, <https://doi.org/10.3189/1985AoG6-1-9-14>, 1985.
- McCormack, E. and Vaa, T.: Testing Unmanned Aircraft for Roadside Snow Avalanche Monitoring, *Transportation Research Record*, 2673, 94–103, <https://doi.org/10.1177/0361198119827935>, 2019.
- Meier, L., Jacquemart, M. F., Blattmann, B., and Arnold, B.: Real-Time Avalanche Detection with Long-Range, Wide-Angle Radars for Road Safety in Zermatt, Switzerland, in: *International Snow Science Workshop Breckenridge, CO, USA, 2016*, pp. 304–308, [https://arc.lib.montana.edu/snow-science/objects/ISSW16\\_014.01.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW16_014.01.pdf), last access: 21.02.2024, 2016.

- Meister, R.: Country-wide Avalanche Warning in Switzerland, International Snow Science Workshop ISSW, Snowbird, Utah, USA, pp. 58–71, 1994.
- Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola, P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfuerscheller, C., Poussin, J., Przyluski, V., Thieken, A. H., and Viavattene, C.: Review article: Assessing the costs of natural hazards – state of the art and knowledge gaps, *Natural Hazards and Earth System Sciences*, 13, 1351–1373, <https://doi.org/10.5194/nhess-13-1351-2013>, 2013.
- Molenaar, M.: An introduction to the theory of spatial object modelling for GIS, *Research Monographs in Geographic Information Systems*, Taylor and Francis, United Kingdom, ISBN 9780748407750, 1998.
- Moner, I., Orgué, S., Gavaldà, J., and Bacardit, M.: How big is big: the results of the avalanche size classification survey, *International Snow Science Workshop ISSW*, [https://arc.lib.montana.edu/snow-science/objects/ISSW13\\_paper\\_P1-05.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW13_paper_P1-05.pdf), last access: 07.06.2024, 2013.
- Müller, K., Techel, F., Mitterer, C., Feistl, T., Sofia, S., Roux, N., Palmgren, P., Bellido, G. M., and Bertrand, L.: The EAWS Matrix, a Look-Up Table for Regional Avalanche Danger Level Assessment, and Its Underlying Concept, in: *International Snow Science Workshop Proceedings 2023*, pp. 540–546, Bend, Oregon, [https://arc.lib.montana.edu/snow-science/objects/ISSW2023\\_05.05.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW2023_05.05.pdf), last access: 16.02.2024, 2023.
- Nöthiger, C. and Elsasser, H.: Natural Hazards and Tourism: New Findings on the European Alps, *Mountain Research and Development*, 24, 24–27, [https://doi.org/10.1659/0276-4741\(2004\)024\[0024:nhatnf\]2.0.co;2](https://doi.org/10.1659/0276-4741(2004)024[0024:nhatnf]2.0.co;2), 2004.
- OpenTopography: Shuttle Radar Topography Mission (SRTM) Global, <https://doi.org/10.5069/G9445JDF>, 2013.
- Ortner, G., Michel, A., Spieler, M. B. A., Christen, M., Bühler, Y., Bründl, M., and Bresch, D. N.: Assessing climate change impacts on snow avalanche hazard, <https://doi.org/10.13140/RG.2.2.21671.32167>, [Preprint], 2023.
- Padilla, R., Netto, S. L., and da Silva, E. A. B.: A Survey on Performance Metrics for Object-Detection Algorithms, *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, <https://doi.org/10.1109/IWSSIP48289.2020.9145130>, 2020.
- Paul, F., Barrand, N. E., Baumann, S., Berthier, E., Bolch, T., Casey, K. A., Frey, H., Joshi, S. P., Konovalov, V., Le Bris, R., Mölg, N., Nosenko, G., Nuth, C., Pope, A., Racoviteanu, A., Rastner, P., Raup, B., Scharrer, K., Steffen, S., and Winsvold, S. H.: On the accuracy of glacier outlines derived from remote-sensing data, *Annals of Glaciology*, 54, 171–182, <https://doi.org/10.3189/2013AoG63A296>, 2013.
- Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., van Herwijnen, A., Olevski, T., Obozinski, G., Pérez-Cruz, F., and Schweizer, J.: Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland, *Natural Hazards and Earth System Sciences*, 22, 2031–2056, <https://doi.org/10.5194/nhess-22-2031-2022>, 2022.

- Pielmeier, C., Zweifel, B., Techel, F., C., M., Grüter, S., and Stucki, T.: Schnee und Lawinen in den Schweizer Alpen. *Hydrologisches Jahr 2022/23*, vol. 145, WSL-Institut für Schnee- und Lawinenforschung SLF; Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL, Davos; Birmensdorf, <https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A36046>, last access: 20.3.2024, 2024.
- Portenier, C., Hüsler, F., Härer, S., and Wunderle, S.: Towards a webcam-based snow cover monitoring network: methodology and evaluation, *The Cryosphere*, 14, 1409–1423, <https://doi.org/10.5194/tc-14-1409-2020>, 2020.
- Prakash, N., Manconi, A., and Loew, S.: A new strategy to map landslides with a generalized convolutional neural network, *Scientific reports*, 11, 1–15, <https://doi.org/10.1038/s41598-021-89015-8>, 2021.
- Produit, T., Ingensand, J., and Milani, G.: QGIS plugin or web app? Lessons learned in the development of a 3D georeferencer, *PeerJ Prepr.*, 4, e2243, <https://doi.org/10.7287/peerj.preprints.2243v2>, 2016.
- Pérez- Guillén, C., Sovilla, B., Suriñach, E., Tapia, M., and Köhler, A.: Deducing avalanche size and flow regimes from seismic measurements, *Cold Regions Science and Technology*, 121, 25–41, <https://doi.org/10.1016/j.coldregions.2015.10.004>, 2016.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2021.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, <https://doi.org/10.48550/arXiv.1506.02640>, 2016.
- Robson, B. A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., and Schaffer, N.: Automated detection of rock glaciers using deep learning and object-based image analysis, *Remote Sensing of Environment*, 250, 112 033, <https://doi.org/10.1016/j.rse.2020.112033>, 2020.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Springer International Publishing, Cham, ISBN 978-3-319-24574-4, 2015.
- Rother, C., Kolmogorov, V., and Blake, A.: "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts, *ACM Transactions on Graphics*, 23, <https://doi.org/10.1145/1015706.1015720>, 2004.
- Rudolf-Miklau, F., Sauer Moser, S., and Mears, A. I., eds.: *The technical avalanche protection handbook*, Wilhelm Ernst & Sohn Verlag für Architektur und technische Wissenschaften, Berlin, Germany, ISBN 978-3-433-03034-9, 2014.
- Salm, B.: A short and personal history of snow avalanche dynamics, *Cold Regions Science and Technology*, 39, 83–92, <https://doi.org/10.1016/j.coldregions.2004.06.004>, 2004.
- Sampl, P. and Zwinger, T.: Avalanche simulation with SAMOS, *Annals of Glaciology*, 38, 393–398, <https://doi.org/10.3189/172756404781814780>, 2004.
- Samtaney, R.: A method to solve interior and exterior camera calibration parameters for image resection, Tech. rep., NASA Technical Reports, <https://ntrs.nasa.gov/api/citations/20000101658/downloads/20000101658.pdf>, last access: 13.03.2024, 1999.

- Schimmel, A., Hübl, J., Koschuch, R., and Reiweger, I.: Automatic detection of avalanches: evaluation of three different approaches, *Natural Hazards*, 87, 83–102, <https://doi.org/10.1007/s11069-017-2754-1>, 2017.
- Schweizer, J., Mitterer, C., Techel, F., Stoffel, A., and Reuter, B.: On the relation between avalanche occurrence and avalanche danger level, *The Cryosphere*, 14, 737–750, <https://doi.org/10.5194/tc-14-737-2020>, 2020.
- Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, <https://doi.org/10.48550/arXiv.1409.1556>, 2015.
- Sinha, S., Giffard-Roisin, S., Karbou, F., Deschatres, M., Karas, A., Eckert, N., Coléou, C., and Monteleoni, C.: Can Avalanche Deposits be Effectively Detected by Deep Learning on Sentinel-1 Satellite SAR Images?, in: *Climate Informatics*, Paris, France, <https://hal.archives-ouvertes.fr/hal-02278230>, last access: 01.06.2023, 2019a.
- Sinha, S., Giffard-Roisin, S., Karbou, F., Deschatres, M., Karas, A., Eckert, N., and Monteleoni, C.: Detecting Avalanche Deposits using Variational Autoencoder on Sentinel-1 Satellite Imagery, in: *NeurIPS 2019 Workshop : Tackling Climate Change with Machine Learning* NeurIPS workshop, Vancouver, Canada, <https://hal.archives-ouvertes.fr/hal-02318407>, last access: 01.06.2023, 2019b.
- SLF: Avalanche types, <https://www.slf.ch/en/avalanches/avalanche-science-and-prevention/avalanche-types/>, last access: 2024-03-11, 2024a.
- SLF: Description of automated stations, <https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/measured-values/description-of-automated-stations/>, last access: 2024-03-12, 2024b.
- Sofiiuk, K., Petrov, I., Barinova, O., and Konushin, A.: f-BRS: Rethinking Backpropagating Refinement for Interactive Segmentation, <https://doi.org/10.48550/arXiv.2001.10331>, 2020.
- Sofiiuk, K., Petrov, I. A., and Konushin, A.: Reviving Iterative Training with Mask Guidance for Interactive Segmentation, <https://doi.org/10.48550/arXiv.2102.06583>, 2021.
- Sovilla, B., Schaer, M., and Rammer, L.: Measurements and analysis of full-scale avalanche impact pressure at the Vallée de la Sionne test site, *Cold Regions Science and Technology*, 51, 122–137, <https://doi.org/10.1016/j.coldregions.2007.05.006>, international Snow Science Workshop (ISSW) 2006, 2008.
- Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., and Kelly, J.: A conceptual model of avalanche hazard, *Natural Hazards*, 90, 663–691, <https://doi.org/10.1007/s11069-017-3070-5>, 2017.
- Steinkogler, W., Olivieri, G., Vezzosi, S., Hendrikx, J., Herwijnen, A. V., and Humstad, T.: Infrasound Detection of Avalanches: operational experience from 28 combined winter seasons and future developments, in: *International Snow Science Workshop Proceedings 2018*, Innsbruck, Austria, pp. 621–626, [https://arc.lib.montana.edu/snow-science/objects/ISSW2018\\_007.8.pdf](https://arc.lib.montana.edu/snow-science/objects/ISSW2018_007.8.pdf), last access: 07.06.2024, 2018.

- Stethem, C., Jamieson, B., Schaerer, P., Liverman, D., Germain, D., and Walker, S.: Snow Avalanche Hazard in Canada – a Review, *Natural Hazards*, 28, 487–515, <https://doi.org/10.1023/A:1022998512227>, 2003.
- Stewart, T. R.: Improving reliability of judgemental forecasts, in: *Principles of forecasting: a handbook for researchers and practitioners*, edited by Armstrong, J., pp. 81–106, Kluwer Academic Publisher, ISBN 978-0-7923-7401-5, [https://doi.org/10.1007/978-0-306-47630-3\\_5](https://doi.org/10.1007/978-0-306-47630-3_5), 2001.
- Sultana, F., Sufian, A., and Dutta, P.: Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey, *CoRR*, abs/2001.04074, <https://doi.org/10.48550/arXiv.2001.04074>, 2020.
- Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W., and Wang, J.: High-Resolution Representations for Labeling Pixels and Regions, *ArXiv*, <https://doi.org/10.48550/ARXIV.1904.04514>, 2019.
- Supervisely: Supervisely Computer Vision platform, <https://supervisely.com>, <https://supervisely.com>, last access: 20.07.2023, 2023.
- swisstopo: swissALTI3D - Das hoch aufgelöste Terrainmodell der Schweiz, <https://www.swisstopo.admin.ch/de/hoehenmodell-swissalti3d>, last access: 13.03.2024, 2018.
- swisstopo: SWISSIMAGE- Das digitale Orthofotomosaik der Schweiz, <https://www.swisstopo.admin.ch/de/orthobilder-swissimage-10-cm>, last access: 13.03.2024, 2020a.
- swisstopo: Swiss Map Raster- Produktdokumentation, <https://www.swisstopo.admin.ch/de/landeskarte-swiss-map-raster-25>, last access: 13.03.2024, 2020b.
- Techel, F.: On consistency and quality in public avalanche forecasting: a data-driven approach to forecast verification and to refining definitions of avalanche danger, Ph.D. thesis, Department of Geography, University of Zurich, Zurich Switzerland, <https://doi.org/10.5167/uzh-199650>, 2020.
- Techel, F., Müller, K., and Schweizer, J.: On the importance of snowpack stability, the frequency distribution of snowpack stability, and avalanche size in assessing the avalanche danger level, *The Cryosphere*, 14, 3503–3521, <https://doi.org/10.5194/tc-14-3503-2020>, 2020.
- Techel, F., Mayer, S., Pérez-Guillén, C., Schmudlach, G., and Winkler, K.: On the correlation between a sub-level qualifier refining the danger level with observations and models relating to the contributing factors of avalanche danger, *Natural Hazards and Earth System Sciences*, 22, 1911–1930, <https://doi.org/10.5194/nhess-22-1911-2022>, 2022.
- Tonnel, M., Wirbel, A., Oesterle, F., and Fischer, J.-T.: *AvaFrame com1DFA (v1.3)*: a thickness-integrated computational avalanche module – theory, numerics, and testing, *Geoscientific Model Development*, 16, 7013–7035, <https://doi.org/10.5194/gmd-16-7013-2023>, 2023.
- Trevethan, R.: Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice, *Frontiers in public health*, 5, 307, <https://doi.org/10.3389/fpubh.2017.00307>, 2017.

- UNISDR, U.: Terminology on disaster risk reduction, Geneva, Switzerland, 2009.
- University of Innsbruck, Lo.La Peak Solutions GmbH, Avalanche Warning Service Tyrol, and Avalanche Warning Service Bavaria: UIBK Avalanche Dataset [data set], <https://doi.org/10.48323/H07F4-QZD17>, 2023.
- U.S. Department of Agriculture: Snow avalanches: a handbook of forecasting and control measures, *Agricultural Handbook*, 194, 84 p., <https://naldc.nal.usda.gov/download/CAT87208423/pdf>, last access: 11.11.2022, 1961.
- Van Coillie, F. M., Gardin, S., Anseel, F., Duyck, W., Verbeke, L. P., and De Wulf, R. R.: Variability of operator performance in remote-sensing image interpretation: the importance of human and external factors, *International Journal of Remote Sensing*, 35, 754–778, <https://doi.org/10.1080/01431161.2013.873152>, 2014.
- Vickers, H., Eckerstorfer, M., Malnes, E., Larsen, Y., and Hindberg, H.: A method for automated snow avalanche debris detection through use of synthetic aperture radar (SAR) imaging, *Earth and Space Science*, 3, 446–462, <https://doi.org/10.1002/2016ea000168>, 2016.
- Voellmy, A.: Über die Zerstörungskraft von Lawinen, *Schweizerische Bauzeitung*, 73, 280–285, <https://doi.org/10.5169/seals-61910>, 1955.
- Völk, M. S.: Analyse der Beziehung zwischen Lawinenauslösung und prognostizierter Lawinengefahr: Quantitative Darstellung einer regionalen Lawinenaktivität am Beispiel Davos (CH), Master thesis, Leopold-Franzens-Universität Innsbruck, 2020.
- Waldeland, A. U., Reksten, J. H., and Salberg, A.-B.: Avalanche Detection in SAR Images Using Deep Learning, in: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2386–2389, <https://doi.org/10.1109/IGARSS.2018.8517536>, 2018.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., and Xiao, B.: Deep High-Resolution Representation Learning for Visual Recognition, <https://doi.org/10.48550/arXiv.1908.07919>, 2020.
- Wang, T., Wu, D. J., Coates, A., and Ng, A. Y.: End-to-end text recognition with convolutional neural networks, *Proceedings - International Conference on Pattern Recognition*, p. 3304 – 3308, <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84874562673&partnerID=40&md5=6d15c26e59d64adf23e0a2d9ac6d5965>, 2012.
- Weller, S. and Mann, N.: Assessing rater performance without a “gold standard” using consensus theory, *Medical Decision Making*, 17, 71 – 79, <https://doi.org/10.1177/0272989X9701700108>, 1997.
- Wesselink, D. S., Malnes, E., Eckerstorfer, M., and Lindenbergh, R. C.: Automatic detection of snow avalanche debris in central Svalbard using C-band SAR data, *Polar Research*, 36, 1333–1336, <https://doi.org/10.1080/17518369.2017.1333236>, 2017.
- Wever, N., Vera Valero, C., and Techel, F.: Coupled snow cover and avalanche dynamics simulations to evaluate wet snow avalanche activity, *Journal of Geophysical Research: Earth Surface*, 123, 1772–1796, <https://doi.org/10.1029/2017JF004515>, 2018.
- Wiesmann, A., Wegmüller, U., Honikel, M., Strozzi, T., and Werner, C. L.: Potential and methodology of satellite based SAR for hazard mapping, *IGARSS 2001. Scanning the Present*



- and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), 7, 3262–3264 vol.7, <https://api.semanticscholar.org/CorpusID:54045258>, last access: 25.03.2024, 2001.
- Wilhelm, C., Wiesinger, T., Bründl, M., and Ammann, W.: The Avalanche Winter 1999 in Switzerland - An Overview, in: Proceedings of the 2000 International Snow Science Workshop, October 1-6, Big Sky, Montana, pp. 487–494, <https://arc.lib.montana.edu/snow-science/objects/issw-2000-487-494.pdf>, 2000.
- Wingtra: Wingtra One, <https://wingtra.com/mapping-drone-wingtraone/>, last access: 13.03.2024, 2024.
- Wong, K., Paritosh, P., and Bollacker, K.: Are ground truth labels reproducible? An empirical study, in: Proceedings of ML Evaluation Standards Workshop at ICLR 2022, pp. 1–12, <https://ml-eval.github.io/assets/pdf/GroundTruthReproducibilityICLRSubmitted.pdf>, 2022.
- WSL Institute for Snow and Avalanche Research SLF (ed.): Avalanche Bulletin Interpretation Guide, WSL Institute for Snow and Avalanche Research SLF, Edition November 2021, 53 p., [https://www.slf.ch/files/user\\_upload/SLF/Lawinenbulletin\\_Schneesituation/Wissen\\_zum\\_Lawinenbulletin/Interpretationshilfe/Interpretationshilfe\\_EN.pdf](https://www.slf.ch/files/user_upload/SLF/Lawinenbulletin_Schneesituation/Wissen_zum_Lawinenbulletin/Interpretationshilfe/Interpretationshilfe_EN.pdf), 2021.
- Xu, N., Price, B., Cohen, S., Yang, J., and Huang, T.: Deep Interactive Object Selection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 373–381, <https://doi.org/10.1109/CVPR.2016.47>, 2016.
- Xu, Z. and Zhao, S.: Fine-grained urban blue-green-gray landscape dataset for 36 Chinese cities based on deep learning network, *Scientific Data*, 11, <https://doi.org/10.1038/s41597-023-02844-2>, 2024.
- Yuan, Y., Chen, X., and Wang, J.: Object-Contextual Representations for Semantic Segmentation, in: *Computer Vision – ECCV 2020*, pp. 173–190, Springer International Publishing, [https://doi.org/10.1007/978-3-030-58539-6\\_11](https://doi.org/10.1007/978-3-030-58539-6_11), 2020.
- Zweifel, B., Lucas, C., Hafner, E. D., Techel, F., Marty, C., and Stucki, T.: Schnee und Lawinen in den Schweizer Alpen. Hydrologisches Jahr 2018/19, vol. 86 of *WSL Berichte*, WSL-Institut für Schnee- und Lawinenforschung SLF; Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL, Davos; Birmensdorf, <https://www.dora.lib4ri.ch/wsl/islandora/object/wsl%3A22232>, last access: 27.02.2024, 2019.