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

Elisabeth D. Hafner[1,2,3], Theodora Kontogianni[3,4], Rodrigo Caye Daudt[3], Lucien Oberson[1,2,5], Jan Dirk Wegner[3,6], Konrad Schindler[3], and Yves Bühler[1,2]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, 7260 Switzerland
[2]Climate Change, Extremes, and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, 7260 Switzerland
[3]EcoVision Lab, Photogrammetry and Remote Sensing, ETH Zurich, Zurich, 8093 Switzerland
[4]ETH AI Center, ETH Zurich, Zurich, 8092 Switzerland
[5]Swiss National Railway, SBB, Berne, 3000 Switzerland
[6]Department of Mathematical Modeling and Machine Learning, University of Zurich, Zurich, 8057 Switzerland

Correspondence: Elisabeth D. Hafner (elisabeth.hafner@slf.ch)

Abstract.

For many safety-related applications such as hazard mapping or road management, well documented avalanche events are crucial. Nowadays, despite research into different directions, the available data is mostly restricted to isolated locations where it is collected by observers in the field. Webcams are getting more frequent in the Alps and beyond, capturing numerous avalanche prone slopes several times a day. To complement the knowledge about avalanche occurrences, we propose to make 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 to support 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 prediction 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. If mounted in a stable position, the camera can be georeferenced with a mono-photogrammetry tool, allowing for exact geolocation of the avalanche outlines and subsequent use in geographical information systems (GIS). In this way all avalanches mapped in an image can be imported into a designated database, making them available for the relevant safety-related applications. For imagery, we rely on current and archive data from webcams that cover the Dischma valley near Davos, Switzerland and capture an image every 30 minutes during daytime since the winter 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.

1 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., 2015). 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 validate and further develop numerical avalanche simulation software like SAMOS or RAMMS (Sampl and Zwinger, 2004; Christen et al., 2010). Today information on occurred avalanches is still mainly reported and collected at isolated locations, unsystematically 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 as well as 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. The good temporal resolution, oftentimes between 10 and 60 minutes, allows for an near-realtime determination of release time. Furthermore, the sequence of images heightens the chance of obtaining an image without low cloud cover or fog that would prevent avalanche documentation of the whole avalanche. Except for our own initial proposition (Hafner et al., 2023) and 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. Opposed to their focus on finding the images and areas containing avalanches, we are aiming at extracting the exact avalanche outlines from the imagery.

There is 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 approx. 2m, or finer; Hafner et al., 2021, 2023), needs to be ordered and captured upon request which is expensive and dependent on cloud free weather conditions. Radar data has 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 approx. every 12 days in Switzerland) and other suitable radar data needs 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, if suitable satellite data is available areas affected by avalanches may be identified and documented continuously over large regions with identical methodology.

Identifying and delineating individual avalanches in any image is a form of instance segmentation, the canonical problem to detect individual objects and determine their outlines. 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 in-
stance 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 predictions, 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 prediction
where the automatic model makes mistakes, gradually refining the segmentation result. The goal is an accurate segmentation
mask, provided by the IOS model with as little user input as possible. The key difference 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 discs of three to five
pixel radius or to Gaussians, depending on the model.

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. Since
most webcams are mounted in a stable position, always capturing the same area, they may be georeferenced for a transfer of the
avalanche area identified in the image to a map. There are several monophotogrammetry tools available to georeference single
images, initially developed to georeference old photographs (e.g., Bozzini et al., 2012, 2013; Produit et al., 2016; Golparvar
and Wang, 2021). Only with existing georeferencing the detected avalanches can be exactly geolocated, compared by size,
aspect or slope angle as well as imported into existing long-term databases.

To complement the currently established ways avalanche occurrences are documented, we propose to make 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) and investigate the transferability of
our model results to the real world application 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 minutes, 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 monophotogrammetry 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.

2 Data

2.1 SLF Webcam network

Almost the whole Dischma Valley, a high alpine side valley of Davos, is covered by our webcams network made up of fourteen
cameras mounted at six different locations (Fig. 1). The valley is about 13km long, the valley floor reaches from 1500m a.s.l to

2000m a.s.l, while the summits reach heights over 3000m a.s.l.. The Dischma valley is permanently inhabited in the lower five kilometers while the road leading to its upper part is closed in winter. With steep mountains on both sides of the valley over 80% of the entire area are potential avalanche terrain (Bühler et al., 2022). Outside the permanent settlements avalanches can 90 only be monitored remotely, especially during high avalanche danger.

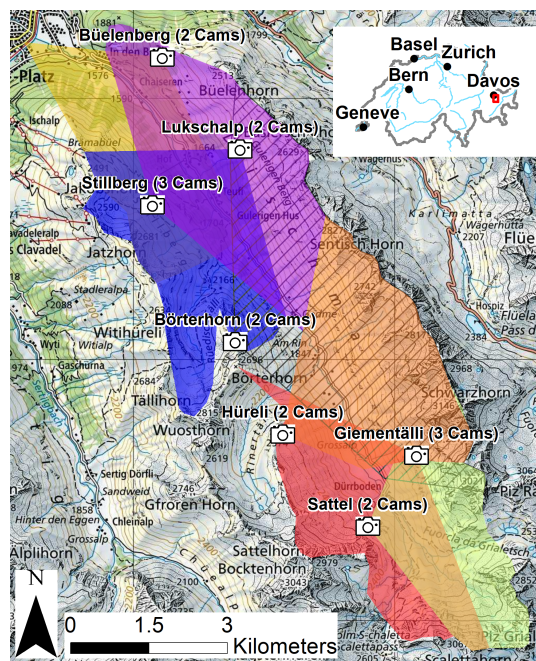



Figure 1. Locations and area covered by the fourteen cameras mounted in six different locations in the Dischma Valley, Davos. The Hüreli station succeeded the Börterhorn station which is no longer in operation (map source: Federal Office of Topography).

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. 2). The acquisition of images every 30 minutes 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 95 they have occurred. The images are then sent to SLF in near-real time via the mobile network and stored on a server. The first camera was mounted at the Büelenberg station in summer 2019, with the next four stations being established in the following months. The Börterhorn station came later, has only been in operation from December 2021 to June 2023 and has been moved to a new location with similar view in December 2023 (Hüreli station). The images have previously been used in the ESA DeFROST Project (ESA, 2020) and in Baumer et al. (2023) .

(a) Station with two cameras, bolted to a rock face at Lukschalp.   (b) Station with (initially) three cameras, mounted on a mast at Sattel.

Figure 2. The stations in the Dischma valley were either mounted on a mast or bolted to rock faces. They host two to three cameras and all infrastructure necessary to ensure power supply as well as data acquisition and transmission.

2.2 Avalanche images and annotations

SLF dataset

We use imagery from the webcams at our stations for training (all except Börterhorn and Hüreli; Sect. 2.1). The images with a size of 6000×4000 pixels are from seven different cameras that captured well identifiable avalanches since being in operation. For training we prepared the images and cropped to 1000×1000 pixels, keeping only the avalanches and their immediate surrounding in the original resolution. For evaluating and for our user study we want to segment all captured avalanches per image, therefore we only resize the images to 3600×2400, 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. 1). About three quarters are used for training, testing and validation while the rest is 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. 1). The *WebNew* contains mostly small avalanches, some of them captured un-

der diffuse illumination conditions, while the *GroundPic* depicts larger avalanches and includes some images of lower quality 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
115 to the dataset, while the rest are also part of the *WebNew* or the *GroundPic*.

Table 1. Overview of the datasets used.

Dataset name	Avalanche annotations		Description
SLF	*train*	200	Webcam imagery and annotations from our test site in Dischma (Fig. 1).
	vali	44	
	test	45	
	WebNew	46	Imagery and annotations from the Börterhorn station (Fig. 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 from Fox et al. (2023).
	vali	382	
	test	867	

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). 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
120 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 their *UIBK dataset* for training and 382 avalanches for validation which we prepared by cropped to 1000×1000 pixels (Tab. 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
125 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. 3). We resized the images larger than 3600×2400 to that size for the evaluation.

Figure 3. Comparing the details in the annotation from one of the *SLF* webcam images (left) to an image from the *UIBK dataset* (right; Fox et al. (2023)).

3 Methodology

3.1 Model architecture

130 We use a state-of-the-art interactive segmentation model introduced by Sofiiuk et al. (2021): The HRNet+OCR (Wang et al., 2020; Yuan et al., 2020) serves as a backbone (core architecture), while the positive and negative clicks are encoded as discs with a fixed radius of enenson et al., 2019). Semantic segmentation backbones usually take only RGB images as input, for interactive segmentation the handling of additional input, in our case encoded user clicks, needs to be carefully implemented (Fig. 4 and Fig. 4). Sofiiuk et al. (2021)'s solution to this is Conv1S: a convolutional block that outputs a tensor of exactly

135 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 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. A combination of random and iterative sampling strategies are employed to generate clicks for training, with masks from previous steps included in the iterative sampling procedure (Fig. 5). Morphological erosion is used to shrink the largest

140 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) or Li et al. (2018).

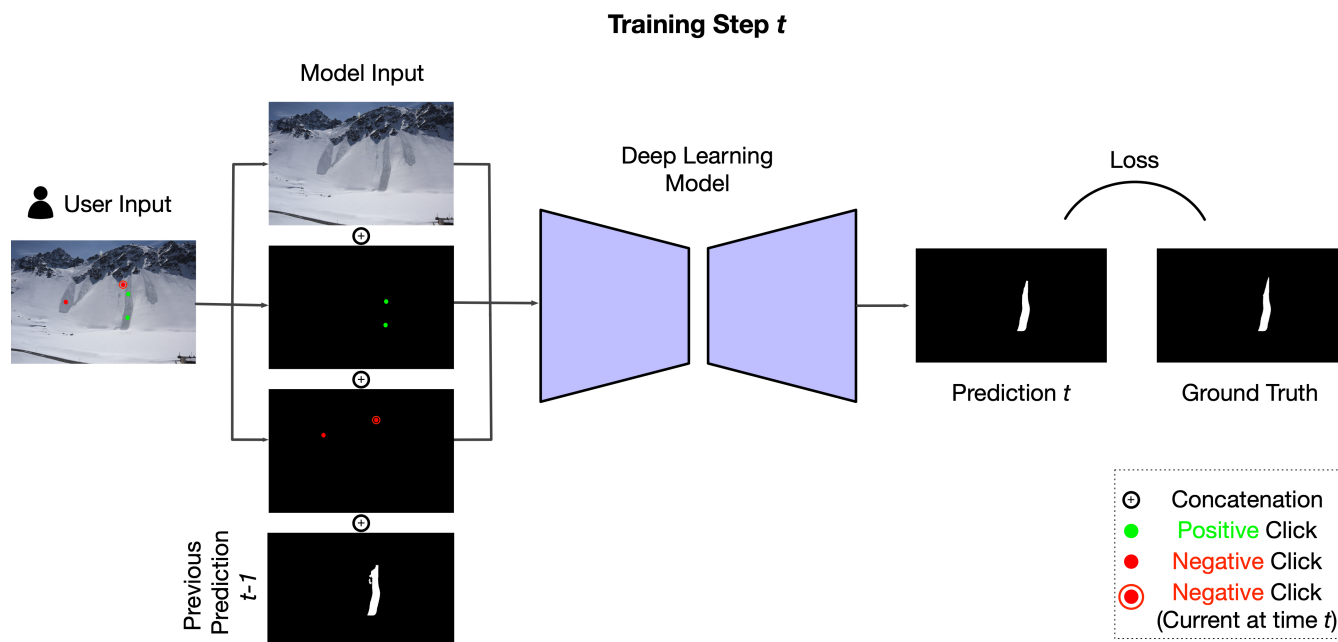



Figure 4. Illustration of the finetuning step of the IOS when training on avalanches.

We make the following adaptions to the original model from Sofiiuk et al. (2021):

145   – we train on patches of 600×600 pixels instead of 320×480, that we crop from varying places of our training images

     – for data augmentation during training we additionally include random translation (max. 3%) and rotation (max. 10 degrees)

     – we replace the manual multistep learning rate scheduler by a cosine learning rate scheduler to profit from a a decreasing learning rate without the need to fiddle with the steps and rates of decay

150   – we do not use the zoom-in function

     – we use a batch size of 4 instead of 28 due to our relatively small training dataset but fine image resolution

3.2   Evaluation metrics

The raw predictions (i.e., the per-pixel probabilities for being part of the avalanche) are thresholded at 0.5 to obtain binary avalanche masks for the analyses. We use the Intersection over Union (IoU) as an indicator of spatial agreement between either

155   the predicted and ground truth masks or the bounding boxes around those masks (e.g., Levandowsky and Winter, 1971).

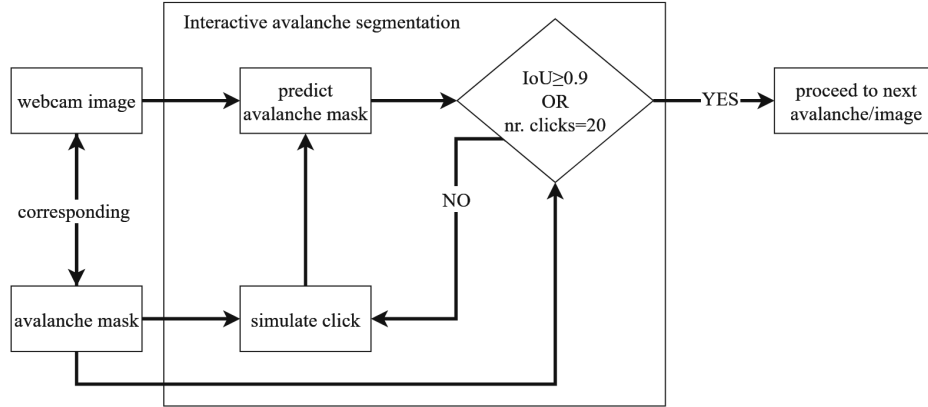



Figure 5. Illustration on the handling of one avalanche when training the IAS model with clicks generated by random and iterative sampling. For the new prediction all previous clicks, as well as the previous mask (if available) are considered.

Pixel-wise metrics

On the pixel level of the masks we report the average Number of Clicks (NoC) necessary to reach IoU thresholds of 0.8 and 0.9, respectively (denoted as mNoC@80 and mNoC@90). Furthermore, we compare the IoU at click k (for $k = 1,2,....,20$) averaged over all the images (mIoU@k), since the we aim for a high IoU with as few clicks as possible. Additionally, we report the number of images that do not reach 0.85 IoU, even after 20 clicks (NoC_{20}@85).

Object-wise metrics

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

From the matches we compute the F1-score as

$$\text{F1} = 2 \cdot \frac{\text{PPV} \cdot \text{POD}}{\text{PPV} + \text{POD}}, \tag{1}$$

where Probability of Detection (POD) and Positive Predictive Value (PPV) are defined as

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

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

3.3 Experimental setup

To find the best model for interactively segmenting avalanches from our webcam imagery we evaluate several training regimes, all with the same model architecture (see Sect. 3.1). Our baseline is a model trained only on *COCO+LVIS* (104k images and

175 1.6M instance-level masks; Lin et al., 2015; Gupta et al., 2019), meaning that it has never seen an avalanche. We then create three further versions by 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. 1). Preliminary tests confirmed that fine-tuning the model pre-trained on *COCO+LVIS* is always superior to training from scratch using only avalanche data. This is in line with previous work on avalanches (Hafner et al., 2022). We perform hyperparameter tuning on the validation set

180 (e.g. selecting the ideal number of training epochs: 90 for AvaWeb and AvaPic, 95 for AvaMix). We use the parameters selected on the validation fixed during our evaluation on the test set. For evaluation, we test how well the model generalizes to the *SLF test* as well as to images from other webcams (*WebNew*). We additionally evaluate on the *GroundPic* and the *UIBK test* to assess the robustness of the model configurations to images from outside our webcam perspective. In addition we compare to segmentation results from previous work by (Fox et al., 2023), by calculating bounding boxes for our predictions and evaluating

185 their overlap with respect to the ground truth bounding boxes from the *UIBK test*.

3.4 User Study

The way click locations are chosen in the model has to be kept simple to reduce computational power. This may however lead to a gap between simulated clicks and real user behaviour. Therefore, it is important to explore if the way the model has learned to make avalanche segmentation faster also applies when real-users click. To investigate if the metrics from evaluating our model hold with real users who's input is noisier and who may adapt to model behaviour, we carried out a small user study. Eight

190 participants were given a short introduction and mapped one avalanche per *UserPic* image. For our user study we used the GUI provided by Sofiiuk et al. (2021), adapting it to save the click coordinates, the time needed per click as well as 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 perform

195 two trial segmentations that are 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 report the mNoC@80 and mNoC@90, compare the mIoU@k, the mean annotation time, the $NoC_{20}@85$ as well as the differences between the best and worst results in terms of mean IoU. To investigate variability in the avalanche areas identified, like in Hafner et al. (2023), we calculate

200 pairwise IoU scores for the final masks from 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 significance level $p \leq 0.05$.

4 Results

Pixel-wise metrics

205 Evaluating on the *SLF test* the AvaWeb is almost 10% better than the others and almost 25% better than the baseline (*COCO+LVIS*; Fig. 6) from click 1. It remains on top but the others catch up by approximately click 16. AvaPic is consistently the worst at high

click numbers and even drops below the baseline. Adding the data from AvaWeb to AvaPic in the AvaMix improves the results, but only gets about halfway to the AvaWeb alone. Compared to the baseline all models trained with avalanches are 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

210   AvaWeb needs the least clicks for reaching the desired IoU thresholds and only for one image never reaches the $NoC_{20}@85$. The AvaPic, never reaches this threshold for five images while this is the case for only two images for the AvaMix and even the baseline reaches an IoU of 85% for more images. For the remaining analyses we are not going to consider the model trained only on *COCO+LVIS* (baseline).

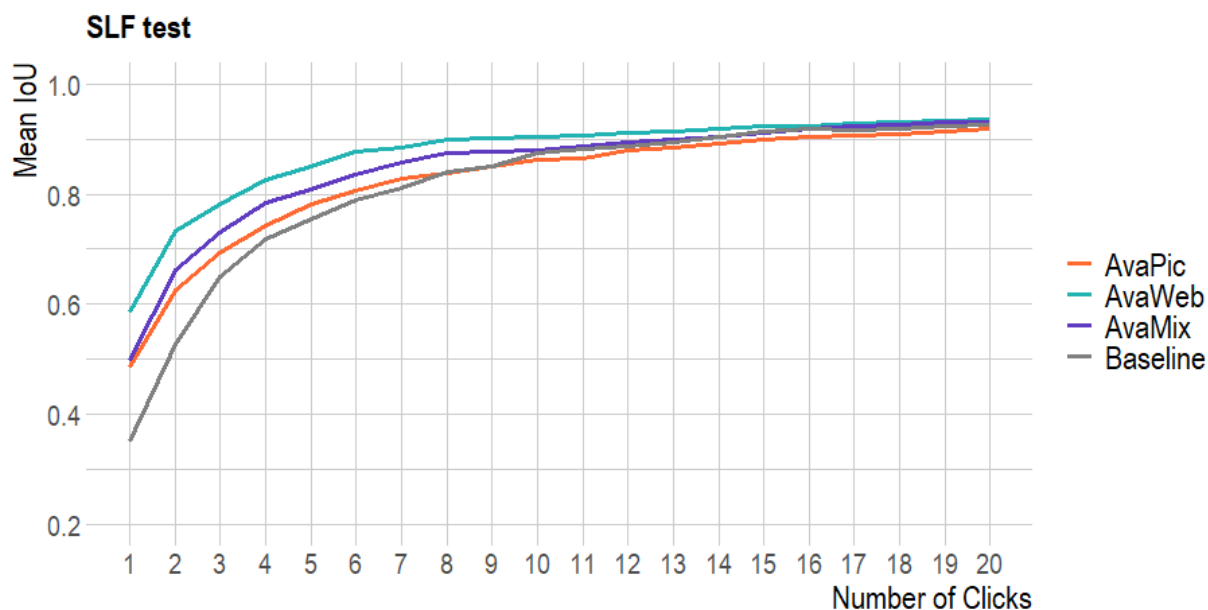



Figure 6. Evaluation on the *SLF test* based on models trained with different datasets: Baseline (*COCO-LVIS*), AvaWeb (*SLF train*), AvaPic (*UIBK train*) and AvaMix (*SLF + UIBK train*).

Table 2. Results for the different datasets when evaluating on the *SLF test*.

Model	Pretrained weights	mIoU@1 [%]	mIoU@2 [%]	mIoU@3 [%]	mNoC@80	mNoC@90	$NoC_{20}@85$
COCO+LVIS (baseline)	-	35.07	52.62	65.00	5.58	9.42	3
AvaWeb	*COCO+LVIS*	**58.59**	**73.40**	**78.30**	**3.31**	**7.6**	**1**
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	9.4	2

To test how well the models generalize to new avalanches under varying perspectives, we test them on the *WebNew*, the
215 *GroundPic* and the *UIBK test* (Fig. 8, Tab. 3): AvaWeb is superior with a margin of up to 30% from click 1 over the AvaPic and
AvaMix on the *WebNew* (Tab. 2; Fig. 7). The AvaPic and AvaMix only catch up around click 10, but never surpass the AvaWeb.
For all models the images in the $NoC_{20}@85$ category depict small, often long and slim avalanches located in the shade, on
imagery acquired under diffuse illumination conditions and/or avalanches that have been snowed on, reducing overall visibility
of the features (Fig. 9).

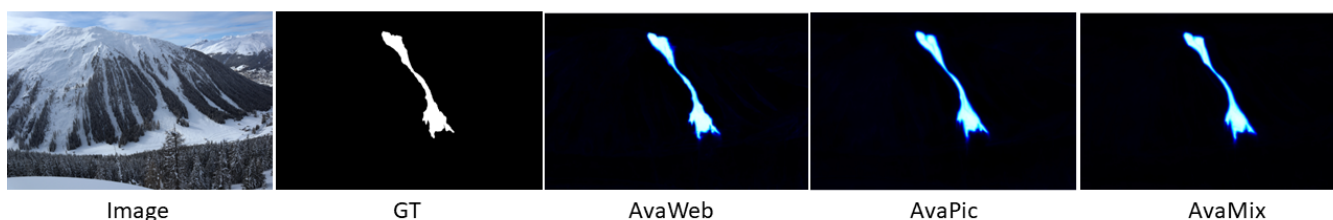



Figure 7. Example for an image from the *SLF test* that all three models solve well. The lighter the hue in the model predictions the higher
the model certainty concerning the existence of an avalanche. In a close-up look the AvaWeb prediction exhibits more nuanced and detailed
avalanche boundaries.

220 On the ground-based *GroundPic* the AvaWeb starts out being the worst by a margin of about 10% while it catches up and
surpasses the AvaPic from click 5 onwards but never reaches the AvaMix. For the large but more coarsely annotated *UIBK
test*, the AvaPic and the AvaMix are consistently superior to the AvaWeb by 10 to 20%. The AvaWeb struggles 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. 10). For some of those avalanches the IoU score reached after
225 20 clicks is well below 50%. Overall for more than one fourth of all avalanches the AvaWeb never reaches the $NoC_{20}@85$,
while this for the AvaPic and AvaMix this is the case for less than 1% of all avalanches. The AvaPic and AvaMix struggle
mostly with the same images which depict 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

Object-wise metrics

230 Comparing bounding boxes the AvaWeb achieves an F1 score 0.13 higher than Fox et al. (2023), from the first click onwards
(threshold 0.1, Tab. 4). For both the AvaPic and the AvaMix the F1 score is even close to 1, therefor by far superior to Fox
et al. (2023) and higher than the AvaWeb. With a threshold of 0.5 for the overlap of the bounding boxes, the AvaPic and the
AvaMix are again superior to the AvaWeb by around 0.2 and remain on top for click 3 and 5 also.

User study

235 For our User Study we loaded 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 was 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

(a) *WebNew*

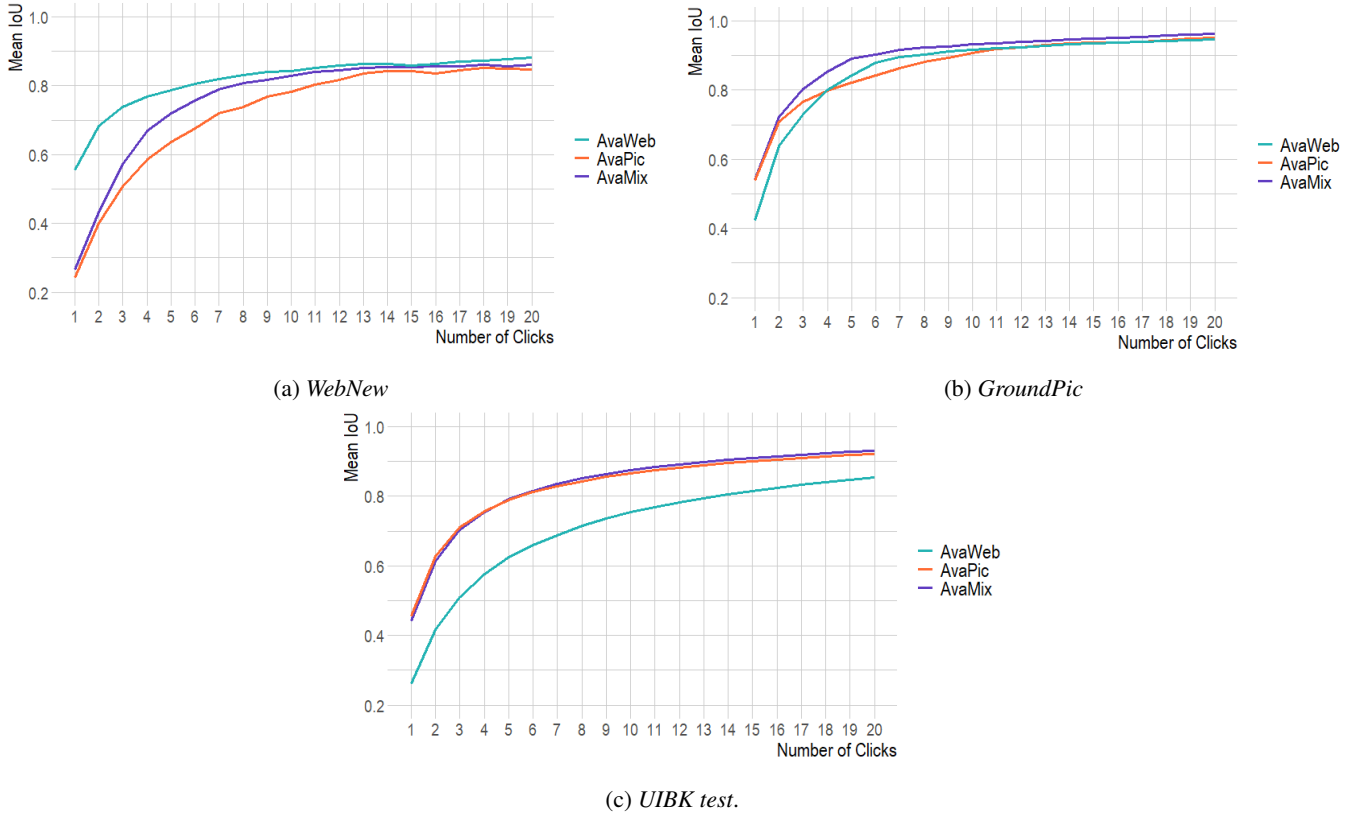
(b) *GroundPic*

(c) *UIBK test*.

Figure 8. Comparing mIoU per click for three datasets with a domain gap to the initial webcam data for our three training configurations: AvaWeb (*SLF train*), AvaPic (*UIBK train*) and AvaMix (*SLF + UIBK train*).

Table 3. Results when testing the generalizability on data not seen during training with a domain gap with respect to the training data.

Dataset	Model	mIoU@1 [%]	mIoU@2 [%]	mIoU@3 [%]	mNoC@80	mNoC@90	NoC_{20}@85
WebNew	AvaWeb	**55.61**	**68.24**	**73.85**	**6.65**	**13.57**	**12**
	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	**6.91**	2
	AvaPic	54.63	71.25	76.92	3.98	7.73	2
	AvaMix	**54.82**	**72.72**	**80.51**	**3.09**	6.96	**1**
UIBK test	AvaWeb	26.19	41.71	51.05	10.47	15.82	246
	AvaPic	44.28	61.29	70.37	**5.84**	**11.26**	**50**
	AvaMix	**45.70**	**62.67**	**70.99**	6.06	11.72	75

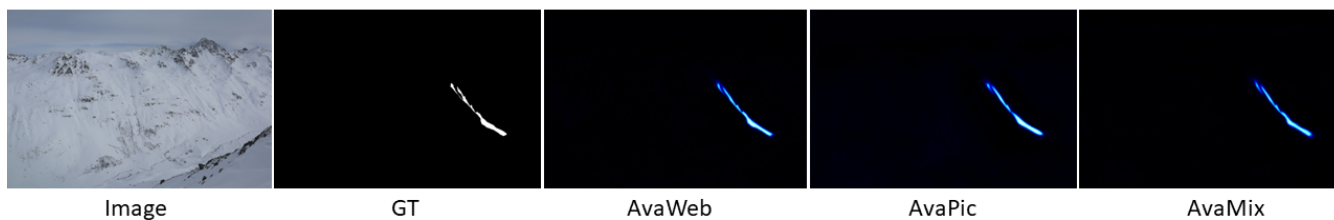



Figure 9. Example for an image from the *WebNew* with diffuse illumination and a long and slim avalanche that all three models struggle with. The lighter the hue in the model predictions the higher the model certainty concerning the existence of an avalanche.

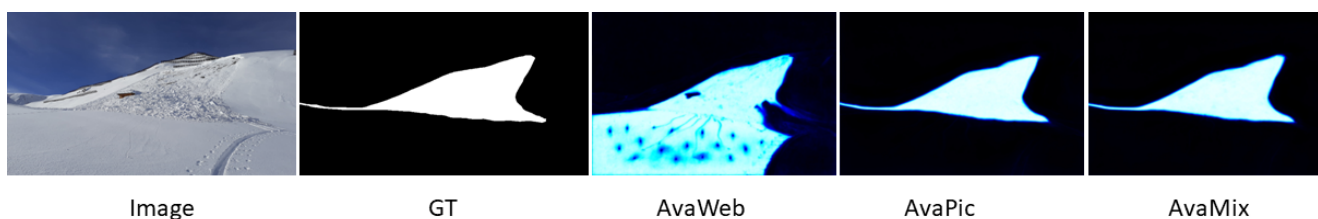



Figure 10. Example of a close-up view of an avalanche from the *GroundPic*, where the AvaWeb struggles with correctly identifying the avalanche area close to the photographer. The lighter the hue in the model predictions the higher the model certainty concerning the existence of an avalanche.

one image remained below that value. This image depicts an avalanche that is located in a partly shaded and partly illuminated
240 area, where especially in the shade features are hard to identify.

On average participants needed 6.5 seconds to reach an IoU of 80% and 9.1 seconds for an IoU of 90%. We do not know how much time is on average spent to map an avalanche with a "traditional method", like the avalanches part of the DAvalMap inventory Hafner et al. (2021). But we had one experienced person record the number of minutes needed for manually mapping about 275 avalanches (size on average 1.75) with the methodology described in Hafner et al. (study 2; 2023): On average 2
245 minutes and 36 seconds were required for mapping one avalanche, time needed ranging from one to eight minutes. This is more than 2 minutes extra than when relying on IAS and translates to a more than 90% saving in time compared to a manual mapping.

In our User study we observed large variations between the different participants: for 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 like in Fig. 11 (top)
250 there is no clear "middle" to place the first click which results in very diverse click strategies for the participants, while for the avalanche in Fig. 11 (bottom) where clicks are placed is more homogeneous- first in the "middle" and then at the top and bottom correcting details. For clicks 1 to 5, where we had enough samples from all participants, we tested if the differences between the highest and the lowest mIoU are statistically significant: While they are not for IoU@1 and IoU@2 (t-test: p-value: > 0.05), for IoU@3 (p-value= 0.045), IoU@4 (p-value= 0.034) and IoU@5 (p-value= 0.035) they are. This is caused by very

Table 4. Comparison of F1 scores and standard deviation for the two different IoU thresholds (5% like Fox et al. (2023), and 50%) on the *UIBK test*.

F1 score ± std	IoU 5%				IoU 50%		
	Fox et al. (2023)	AvaWeb	AvaPic	AvaMix	AvaWeb	AvaPic	AvaMix
automated	0.64 ± 0.60	-	-	-	-	-	-
click 1	-	0.76 ± 0.43	**0.97** ± 0.16	0.96 ± 0.20	0.23 ± 0.42	0.44 ± 0.50	0.42 ± 0.49
click 3	-	0.99 ± 0.11	**1**	1 ± 0.05	0.66 ± 0.47	0.86 ± 0.34	0.87 ± 0.31
click 5	-	1 ± 0.08	**1**	**1**	0.80 ± 0.40	0.94 ± 0.24	0.96 ± 0.20

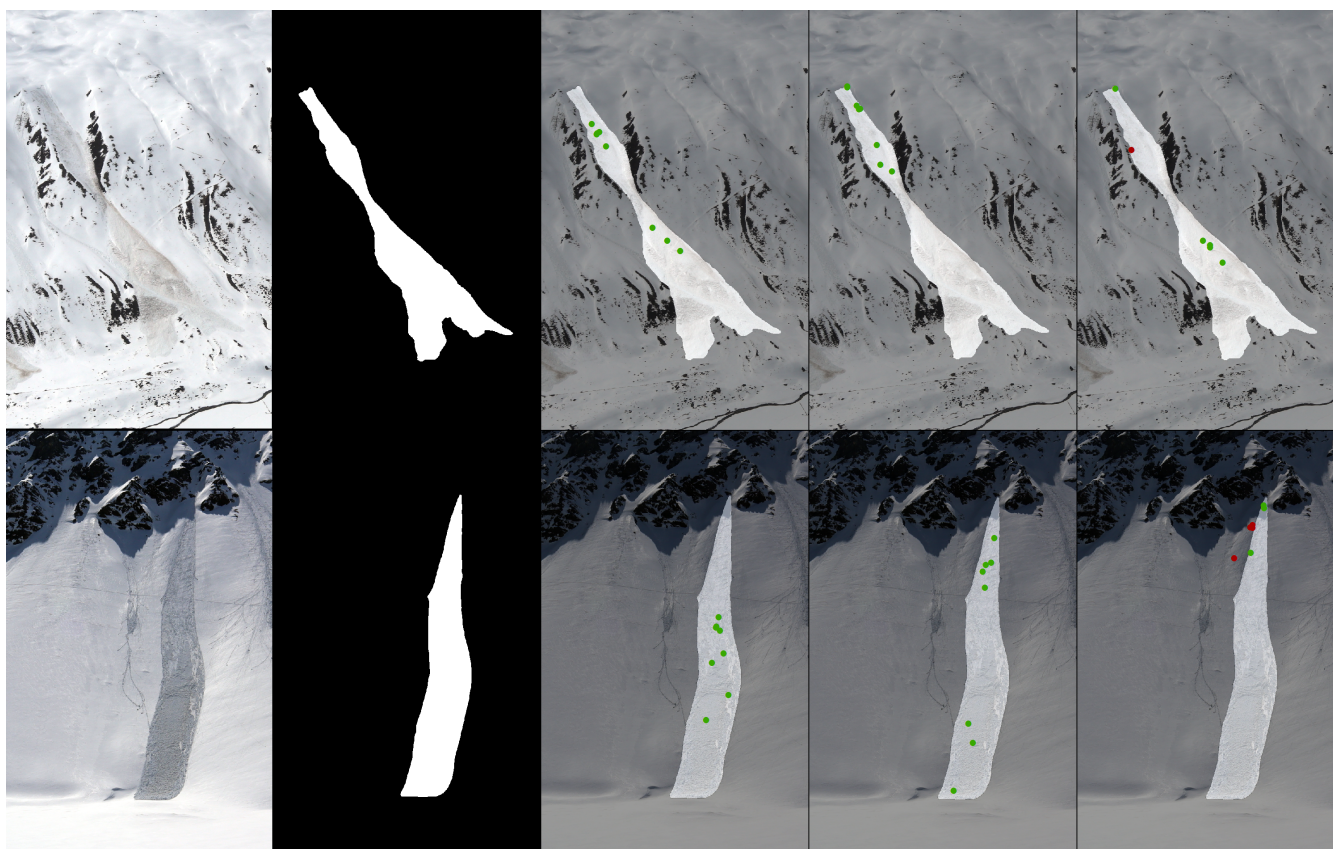



Figure 11. Illustration where the first three clicks in two images from the *UserPic* dataset are placed. Green dots denote positive clicks, red dots denote negative clicks.

255 consistent results with low standard deviation for the participants with the highest mIoU@k scores. When taking the mask from the last click as a final result, the differences between participants are however quite small: the mean pairwise IoU is

93.53%, the maximum 95.44% and the minimum 90.59%. Consequently, all pairs have an IoU within 5% of eachother as their segmented final avalanche masks are very similar (Fig. 13).

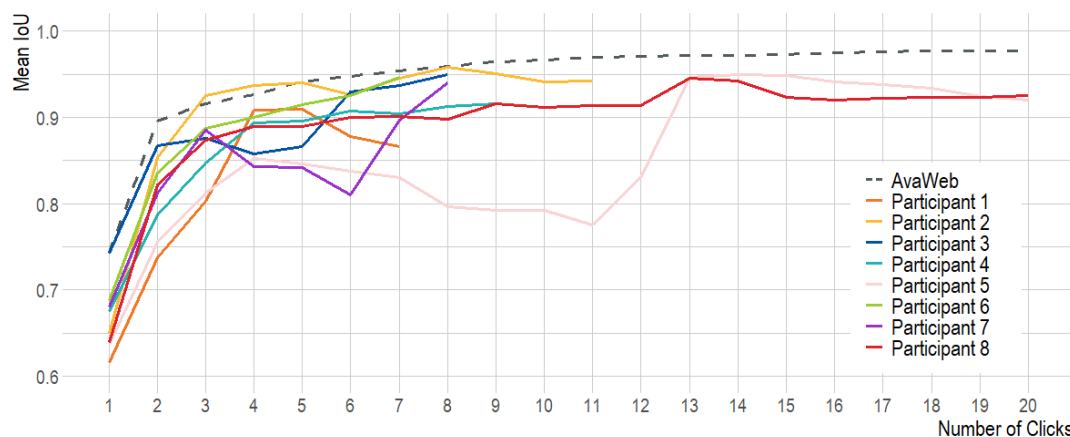



Figure 12. Comparison of the mIoU for all participants of the User study to the mIoU of the AvaWeb evaluated on the *UserPic* dataset. Note that only two participants used the maximum possible number of 20 clicks.

When evaluating AvaWeb on the *UserPic* with simulated clicks and comparing to the User study results (see Tab. 5), the AvaWeb results are superior for all investigated metrics, except the mNoC@80. The participants with the highest mIoU@k hold up to the numbers from the model (Fig. 13).

Table 5. Comparison of the results from the User study with the model results when evaluating on the same imagery (*UserPi*

	User Study	AvaWeb
mNoC@80	2.11	1.85
mNoC@90	2.50	2.55
NoC_{20}@85	1	0
mIoU@1 [%]	66.61	74.31
mIoU@2 [%]	80.91	89.57
mIoU@3 [%]	86.22	91.53

5 Discussion

Our results show that IAS enables segmentation of avalanche outlines from webcam imagery within seconds. The AvaWeb performs best for the two test datasets containing webcam imagery (*SLF test* and *WebNew*), performs on par with the dataset with a perspective unlike those of the webcams (*GroundPic*) but fails to generalize well to the large but coarsely annotated *UIBK test* with a large variety of perspectives and resolutions. In contrast, the models trained on larger and more diverse

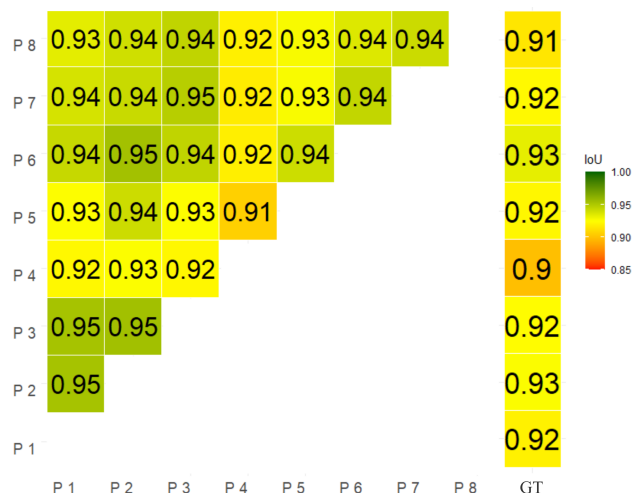



Figure 13. IoU for all participant pairs (Participants denoted as P, the ground truth as GT) for the final masks from our User study on the *UserPic*.

datasets (AvaPic and AvaMix), exhibit lower mIoU scores and a higher amount of clicks to reach a certain IoU for all test sets containing webcam imagery (*SLF test* and *WebNew*), but they perform better on imagery not from webcams (*GroundPic* and *UIBK test*). The AvaMix seems to have learned more details since the mIoU scores are higher than in 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* help the AvaMix to outperform the AvaPic.

Overall, the models struggle 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, 2023). Furthermore, especially the AvaWeb struggles with close-up views of avalanches, oftentimes 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 models, achieves the best performance for two out of three test sets with detailed avalanche annotations (*SLF test*, *WebNew*, *GroundPic*). Even though the *UIBK test* contains 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. But results also show that any model trained on avalanches is better than the baseline which has never before seen an avalanche. Investigating this in more detail is beyond the scope of this paper but for future work we recommend experimenting with a larger dataset of finely annotated avalanches from a variety of perspectives.

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 models bounding boxes on the first click to their results, we

outperform their F1 score by a large margin (0.64 vs. 0.97). Consequently, we capture the area the avalanche covers better from the first prediction onwards.

In our user study the participants with the best performance are as good as the simulation, but the mean IoU scores of all participants cannot beat the model (Tab. 5). We attribute this to the lack of serious training (visible in the variations of the number of clicks and time used) and knowing that estimations of avalanche area exhibit large variabilities (Hafner et al., 2023) as there is no clear unambiguous definition of an avalanche boundary. Since the differences between the model and the participants are rather small, we consider the way user clicks are simulated during training representative of employed real-life click strategies.

Compared to manual mapping using IAS saves about 90% of time needed for mapping, even when compared to manually mapping relatively small avalanches (average size 1.75) that take less time to map 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 is detailed enough. Consequently, the final masks are the most important. As opposed to Hafner et al. (2023) the mean pairwise IoU scores for the avalanche area mapped (pixels in our case), are within 5% of each other and all have an IoU above 0.9 with respect to the ground truth mask (Fig. 13). Consequently, IAS not only improves efficiency but enhances the reliability, defined as the consistency of repeated measurements or judgements of the same event relying on the same process (Cronbach, 1947), as it guides the participants and constrains the results. Even though we had no overlapping avalanches in our *UserPic*, we still believe our findings also apply in this more challenging scenario.

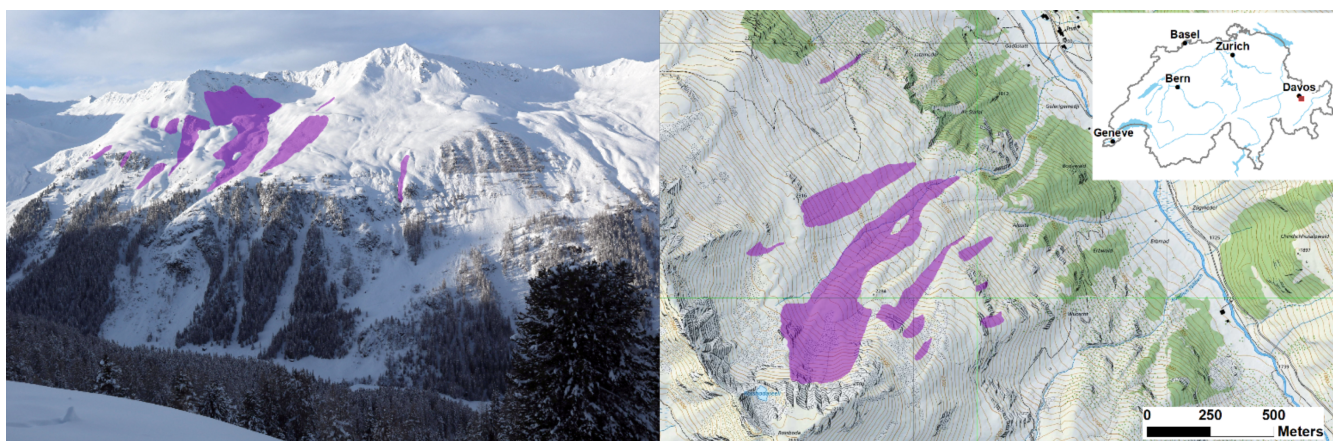



Figure 14. Example of avalanches segmented from an image with AvaWeb (left) and the corresponding avalanches displayed on a map after they have been georeferenced with the monoplotting tool (right, Bozzini et al. (2012); map source: Federal Office of Topography).

Webcams have limited coverage and cannot record avalanches in a spatially continuous manner like satellite imagery may (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 allowing for more detailed analysis of e.g. predisposition factors. The inclusion into existing databases however requires, the georeferenc-

ing of the avalanches, achievable with e.g. monophotogrammetry tools like 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 (like exemplary shown in Fig. 14). Without that the application is limited to providing an overview on the current activity to
310 an avalanche warning service while all other downstream applications cannot profit from the data. As long as the camera is not moved and the image section remains stable, the georeferencing needs to be done only once per camera and can be reused for all subsequent images.

As opposed to fully automatic avalanche segmentation IAS requires a human annotator. We do not see this as a disadvantage, but rather complementary since humans are present and will remain present in the future in many settings where avalanches
315 are recorded, either connected to work or as part of winter leisure activities in the mountains.

Compared to the traditional way of mapping avalanches IAS saves over 90% time, even though we believe that we underestimate the average manual mapping time per avalanche since the avalanches were time was recorded were rather small (mean size 1.7 ha and all located in an area well known to the one person mapping.

6 Conclusions and Outlook

320 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 predictions are 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 the better temporal resolution of approximately 10 to 60 minutes increases the likelihood of capturing avalanches
325 even under adverse visibility conditions, offering a very valuable complementary data source for existing avalanche databases. This allows documentation of the avalanche activity for a whole season compared to just one extreme event like in Bühler et al. (2019). Additionally, the release time may be determined with less uncertainty, helping the avalanche warning and research to better connect the snow and weather conditions to avalanche releases.

In combination, IAS and georeferencing have great potential to improve avalanche mapping: Existing monophotogrammetry
330 tools may be used to import avalanches detected with IAS from webcams. Assuming the camera position and area captured is stable the georeferencing can be reused for all subsequent images, like done before 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 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 orien-
335 tation 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 geolocation of avalanches mapped in the field more reliable compared to the "traditional" mapping approach

described in Hafner et al. (2023). The possibility to record observed avalanches in an easy way could also help to motivate
340 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 et al., 2023), segmenting an avalanche with IAS
saves over 90% time and the results are more reliable in terms of consistency between mappings from different individuals.
The model as is may also be used to annotate images or correct existing annotations with minimum user input. These may be
used to develop and enhance models for automatic avalanche segmentation, saving time while generating outlines that follow
345 the visible avalanche textures, easing the learning, thereby getting more accurate and reliable 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.

Code and data availability. The code will be made available upon publication of this paper together with the images and avalanche annota-
tions.

350 *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 manuscript and all co-authors critically reviewed and complemented it.

Competing interests. The authors declare they have no competing interests.

355 *Financial support.* The initial development and mounting of the majority of the webcams was part of the DeFROST project financed by the
European Space Agency (ESA; N.4000127451/19/NL/CLP).

Acknowledgements. We thank the SLF Workshop and Electronics for developing, building and setting up our camera system in the Dischma
valley. We are grateful to Simon Aeschbacher, Jor Fergus Dal, Amelie Fees, Julia Glaus, Matthias Lichtenegger, Isabelle Rittmeyer, Pia
Ruttner-Jansen and Linda Zaugg-Ettlin for participating in our user study. We thank Luis Scherer for recording the time spent on mapping
360 avalanches in the "traditional way".

References

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.

365 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.

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.

370 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.

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-

375 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.

Bründl, M. and Margreth, S.: Integrative Risk Management, in: W. Haeberli & C. Whiteman (Eds.), Snow and Ice-Related Hazards, Risks

380 and Disasters 2015, pp. 263–301, https://doi.org/10.1016/B978-0-12-394849-6.00009-3, 2015.

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,

385 1825–1843, https://doi.org/10.5194/nhess-22-1825-2022, 2022.

Chen, L., Strauch, M., and Merhof, D.: Instance Segmentation of Biomedical Images with an Object-aware Embedding Learned with Local Constraints, 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.

390 Cronbach, L. J.: Test "reliability": Its meaning and determination, Psychometrika, 12, 1–16, https://doi.org/10.1007/bf02289289, 1947.

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.

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.

395 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.

ESA: DEFROST, https://business.esa.int/projects/defrost, last access 08.02.2024, 2020.

Fox, J., Siebenbrunner, A., Reitinger, 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, 2023.

400   Golparvar, B. and Wang, R.-Q.: AI-supported Framework of Semi-Automatic Monoplotting for Monocular Oblique Visual Data Analysis, 2021.

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.

405   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., 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
410   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., Caye Daudt, R., Wegner, J. D., Schindler, K., and Bühler, Y.: Using interactive object segmentation to derive avalanche outlines from webcam imagery, in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, pp. EGU–10 867, https://doi.org/10.5194/egusphere-egu23-10867, 2023.

415   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.

He, K., Gkioxari, G., Dollár, P., and Girshick, R.: Mask R-CNN, https://doi.org/10.48550/arXiv.1703.06870, 2018.

Kapper, K. L., Goelles, T., Muckenhuber, S., Trügler, A., Abermann, J., Schlager, B., Gaisberger, C., Eckerstorfer, M., Grahn, J., Malnes, E.,
420   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.

425   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.

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.

Kontogianni, T., Gygli, M., Uijlings, J., and Ferrari, V.: Continuous Adaptation for Interactive Object Segmentation by Learning from
430   Corrections, 2020.

Levandowsky, M. and Winter, D.: Distance between sets, Nature, 234, 34–35, 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
435   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.

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.

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://api.semanticscholar.org/CorpusID:220734135, 2020.

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.

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://api.semanticscholar.org/CorpusID:21774865, 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.

Rother, C., Kolmogorov, V., and Blake, A.: "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts, 23, https://doi.org/10.1145/1015706.1015720, 2004.

Rudolf-Miklau, F., Sauermoser, S., and Mears, A., eds.: The technical avalanche protection handbook, Ernst & Sohn, Berlin, 2015.

Sampl, P. and Zwinger, T.: Avalanche simulation with SAMOS, Annals of Glaciology, 38, 393–398, https://doi.org/10.3189/172756404781814780, 2004.

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.

Supervisely: Supervisely Computer Vision platform, https://supervisely.com, https://supervisely.com, last access 2023-07-20, 2023.

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, 2020.

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 236, https://doi.org/10.1080/17518369.2017.1333236, 2017.

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.

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, 2020.