



Fully automated snow depth measurements from time-lapse images applying a convolutional neural network

Matthias Kopp, Ye Tuo ^{*}, Markus Disse

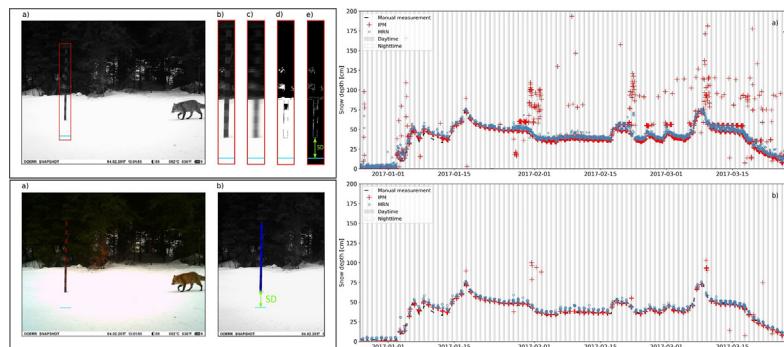
Chair of Hydrology and River Basin Management, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany



HIGHLIGHTS

- First application of Mask R-CNN in snow research.
- Mask R-CNN consistently produced reliable snow depth data.
- Image processing demands specific parameterization for each case.
- Mask R-CNN is a feasible tool for collecting snow depth time series.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 23 April 2019

Received in revised form 29 August 2019

Accepted 30 August 2019

Available online 31 August 2019

Editor: Ralf Ludwig

Keywords:

Mask R-CNN

Instance segmentation

Snow depth

Image processing

Time-lapse camera

ABSTRACT

Time-lapse cameras in combination with simple measuring rods can form a highly reliable low-cost sensor network monitoring snow depth in a high spatial and temporal resolution. Depending on the number of cameras and the temporal recording resolution, such a network produces large sets of image time series. In order to extract the snow depth time series from these collections of images in acceptable time, automated processing methods have to be applied. Besides classic image processing based on edge detection methods, there are nowadays ready-to-use convolutional neural network frameworks like Mask R-CNN that facilitate instance segmentation and thus allow for fully automated snow depth measurements from images using a detectable measuring rod. This study investigates the applicability of Mask R-CNN embedded in a newly developed work flow for snow depth measurements. The new method is compared to an automated image processing method carried out utilizing functionalities provided by the OpenCV library. The quality of both methods was assessed with the inclusion of manual evaluations of the image series. As a result, the newly introduced work flow outperforms the present classic image processing method in regard to stability, accuracy and portability. By applying the Mask R-CNN framework, the overall RMSE of two considered time series is reduced to approximately 20% of the value produced by means of the classic image processing approach. Moreover, the ratio of values within five centimeter deviation from the reference value was increased from 75% to 88% on average. Since no parameters have to be adjusted, the Mask R-CNN framework is able to detect known shapes reliably in almost any environment, making the presented method highly flexible.

© 2019 Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail address: ye.tuo@tum.de (Y. Tuo).

1. Introduction

Snow processes are key components of the hydrological cycle in snow covered regions and play an important role in the seasonal distribution of water resources (Farinotti et al., 2010; Horton et al., 2006; Shinohara et al., 2009; Shrestha et al., 2004; Warscher et al., 2013; Zierl and Bugmann, 2005). Reliable snow information such as snow depth data is not only important for assessing accurate water resource potentials for snow covered districts (Boniface et al., 2015; Harder et al., 2016; Kinar and Pomeroy, 2015; Prokop et al., 2008; Tuo et al., 2018b), but also provides essential indicators that help in the investigation of the impact of climate change (Bell et al., 2016; Szczypta et al., 2015; Xuejin et al., 2018) and in the simulation of floods due to snowmelt by means of a modeling approach (Lawrence et al., 2014; Vormoor et al., 2016).

Many methods have been developed to measure snow depth. Briefly, they can be classified into three groups: 1) classic in-situ measurements by snow probe or rod (measure snow depth by sticking a scaled rod into the snowpack) (Kinar and Pomeroy, 2015; Parajka et al., 2012; Prokop et al., 2008); 2) estimates by sensors such as radar (e.g. ground penetrating radar (GPR)) (McGrath et al., 2015), frequency-modulated continuous wave (FM-CW) (Kwok and Kacimi, 2018), ultrasonic device (Ryan et al., 2008; Schattan et al., 2017), laser device (e.g. light detection and ranging (Lidar)) (Deems et al., 2006; Deems et al., 2013); terrestrial laser scanner (TLS) (Prokop, 2008), neutron device (Schattan et al., 2017), and magnet-based probe (Sturm and Holmgren, 2018)); 3) remote sensed data from satellites (e.g. NOAA CRN (National Oceanic and Atmospheric Administration Climate Reference Network) (Rasmussen et al., 2012), GPS interferometric reflectometry (GPS-IR) (Boniface et al., 2015; Gutmann et al., 2012), and Special Sensor Microwave Imager (SSM/I) (Chang et al., 2005)). The latter two groups were developed to improve the efficiency of snow depth measurements, producing continuous snow depth estimates for larger spatial scales with high efficiency (Boniface et al., 2015; Harder et al., 2016; Prokop et al., 2008; Schattan et al., 2017). A recent comprehensive survey by Pirazzini et al. (2018) about in-situ snow measurements carried out with the data from 99 European institutions shows that snow rod or probe were still the most favored method for measuring snow depth, which accounted for 46% of all the methods. Time-lapse cameras were frequently utilized and made up for 7% of all the methods (Pirazzini et al., 2018). For the rest different sensor based techniques have been applied (Pirazzini et al., 2018). However, sensor based and satellite snow depth datasets are not yet capable to completely replace the classic in-situ measurements due to their limited accuracy (Harder et al., 2016; Prokop et al., 2008) and high cost for installation and maintenance in both money and time. Classic in-situ snow depth measurement is still popular and vital (Pirazzini et al., 2018; Kinar and Pomeroy, 2015; Leppänen et al., 2016): on the one hand, it is commonly recognized as the most accurate dataset with low cost and still essential for calibration and error evaluation of snow depth estimates from sensors and remote sensing techniques (Gutmann et al., 2012; McGrath et al., 2015; Prokop et al., 2008; Raleigh et al., 2015); on the other hand, it is also the most reliable reference for implementing modeling practices (Raleigh et al., 2015; Tuo et al., 2018a) and accurately investigating specific snow characteristics or processes (Leppänen et al., 2016).

One obvious disadvantage of in-situ snow probe or rod measurement is their high demand of manpower for accessing the measurement sites in the field and recording the data, which is of low efficiency and not feasible for collecting continuous data in snow dominated remote areas. To improve the deficiencies, in combination with classic in-situ measurements, time-lapse cameras are powerful tools to record the snow information without the need of field access (Dong and Menzel, 2017; Farinotti et al., 2010; Garvelmann et al., 2013; Parajka et al., 2012; Revuelto et al., 2016) in order to automatize the measurements. The snow depth is subsequently extracted from the images using an

image processing algorithm, which improves the efficiency of classic in-situ snow depth measurement to some extent (Dong and Menzel, 2017; Garvelmann et al., 2013; Parajka et al., 2012). However, the results engendered by the image processing algorithm frequently turn out to be erroneous, which calls for improvement of accuracy, especially when compared to manual recording.

Mask R-CNN, developed recently by He et al. (2017), is a neural network based method. It is a powerful object segmentation framework that allows for accurate pixel-wise instance segmentation in complex environments and has been applied in some disciplines, for example to measure ice wedge polygons (Abolt et al., 2019; Zhang et al., 2018) and to detect objects in transport systems with high speed and high accuracy (https://github.com/matterport/Mask_RCNN (He et al., 2017)). Therefore, if applied properly, Mask R-CNN could improve the accuracy of in-situ snow depth measurements.

In this work, Mask R-CNN is applied to the environment field to produce snow depth time series based on in-situ snow rod photo series for the first time. The results are subsequently compared with those of the image processing method of the same image series. Manual measurements read by eye are recorded and utilized as references for evaluating the two automated methods. By comparing the Mask R-CNN work flow and the image processing method with manual datasets, the specific objectives of this study are to: 1) investigate the accuracy of the two automated methods; 2) evaluate the error frequency derived from the automated methods themselves; 3) compare performances of the two automated methods facing different natural situations (e.g. 24-h light changes; variable light conditions due to topography and extreme weather); 4) assess the improvement of datasets by Mask R-CNN in terms of accuracy, error frequency, adaptability and computation time. The image processing method and the Mask R-CNN workflow used in this study are hereinafter referred to as IPM and MRN, respectively.

2. Material and method

2.1. Study area and instrument set-up

The image series used in this study have been recorded in the experimental, montane catchment of the Dreisäulerbach in the Bavarian Alps (see Fig. 1a). The catchment, which is part of the Isar river system, covers approximately 1.4 km² and is located in the Ammergauer Alps in direct proximity of the Linderhof castle about 120 km south of Munich. The conjunction of the Dreisäulerbach and the Linder streams lies about 940 m above sea level. At an altitude of 1020 m, the outflow from the experimental catchment is recorded by a measuring weir. The highest point of the catchment lies >1600 m above sea level. The area is mostly made up by south facing slopes, but also contains northern slopes in southern parts of the catchment. Apart from small clearings in the lower regions of the area, the catchment is densely forested by spruce and beech. In the higher reaches of the catchment, typical sub-alpine spruce forests with larger stretches of meadows are found. According to the German Weather Service, the mean annual air temperature in the nearby city of Garmisch-Partenkirchen is 7.2 °C and the long-term mean annual precipitation at the Ettal-Linderhof station of the Water Science Service Bavaria is reported to be 1676 mm. Due to its high variability, the catchment was chosen to host an experimental measurement network in order to investigate snow processes in montane and sub-alpine areas of the Bavarian Alps within the frame of the ProMoS research project launched by the Bavarian State Ministry of the Environment and Consumer Protection. In order to monitor snow cover development during the winter, several camera-based snow depth measurement locations are positioned in the experimental catchment. We have chosen two image series for this study: Images from the winter of 2016/17 from the "Rosseck" station and a photo series from the "Talschluss" station from the winter of 2017/18. Considering two time series allows for cross validation of the two investigated measurement methods. The two measurement sites differ in

surrounding vegetation as well as geographical orientation: While the camera set-up at the "Rosseck" station is facing south and is enclosed by coniferous forests, the camera at the "Talschluss" station is directed to the east and the vegetation in the vicinity of the measurement is sparse. These differences in measurement set-up conditions result in different light situations affecting the quality of the recorded image series and the behavior of the two investigated automated measurement methods.

The image time series were taken using a "Dörr Snapshot Mini" time-lapse camera (Fig. 1b). This model is equipped with an infrared flash light that enables night vision for a distance of up to 15 m. The camera is powered by eight AA batteries and an additional external 6 V battery block. Due to its screw thread, the camera can be mounted almost anywhere using an angle joint. In our measurement set-up, the cameras are mounted to a tree trunk (Fig. 1c) and a

wooden pole respectively. The utilized measuring rods are simple battens showing alternating colors every ten centimeters (Fig. 1d).

2.2. Description of IPM

As described above, we aim to introduce a new method of automated snow depth measurements utilizing the MRN framework, an instance segmentation method based on a fully convolutional neural network. On the one hand, the quality and accuracy of the method were evaluated by comparison with computer aided manual measurements; on the other hand the method was compared with an IPM algorithm. The different filtering steps of the IPM procedure using convolution kernels (see Shapiro & Stockman, 2001), as depicted in Fig. 2a, were performed with the "OpenCV" open source software package (Bradski, 2000), which is available for Python, C++ and Java. Prior

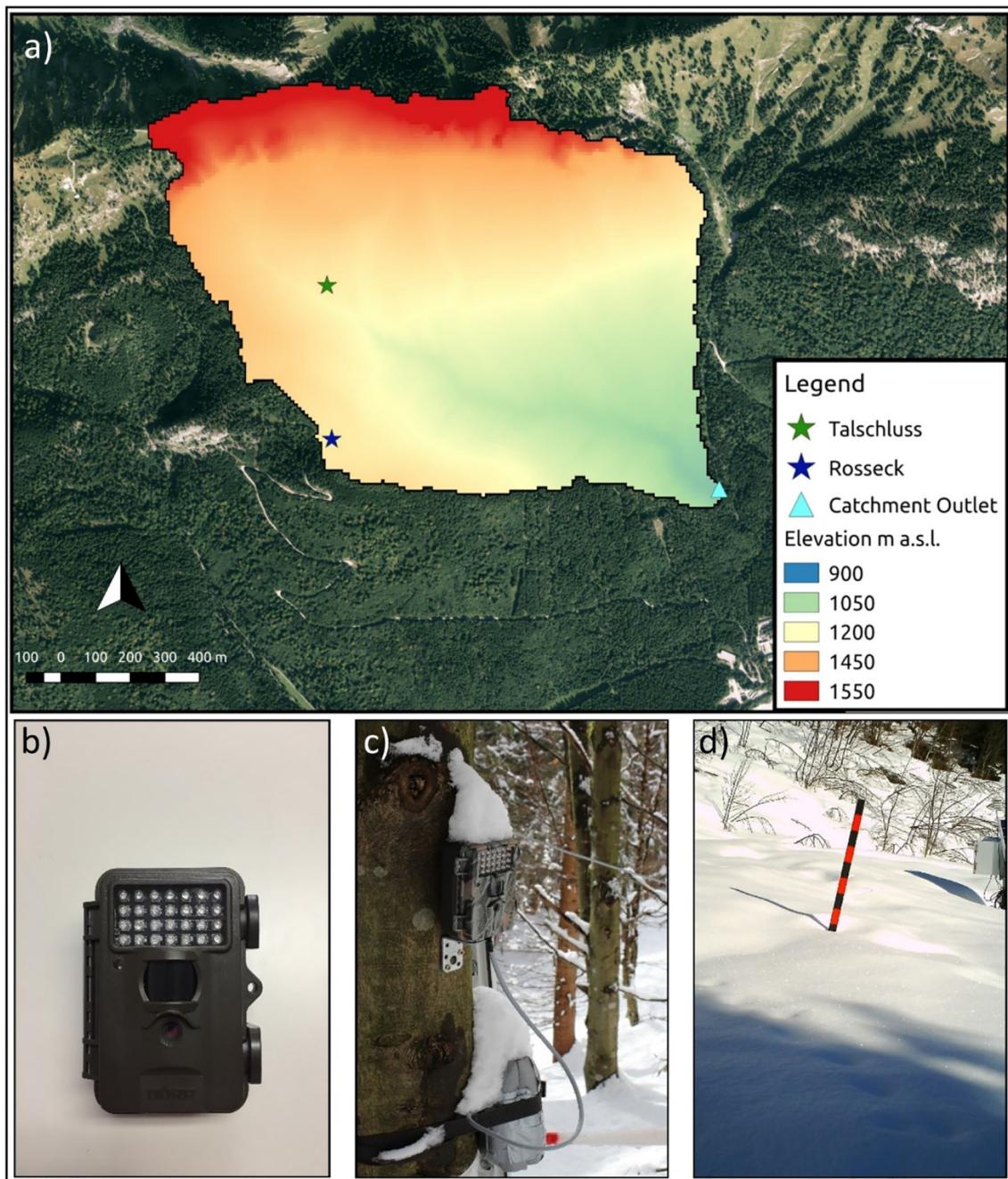


Fig. 1. Study area and instrument set-up of the snow measurements. a) Dreisäulerbach and the measurement locations; b) time-lapse camera, c) camera mounted to tree, d) sample photo

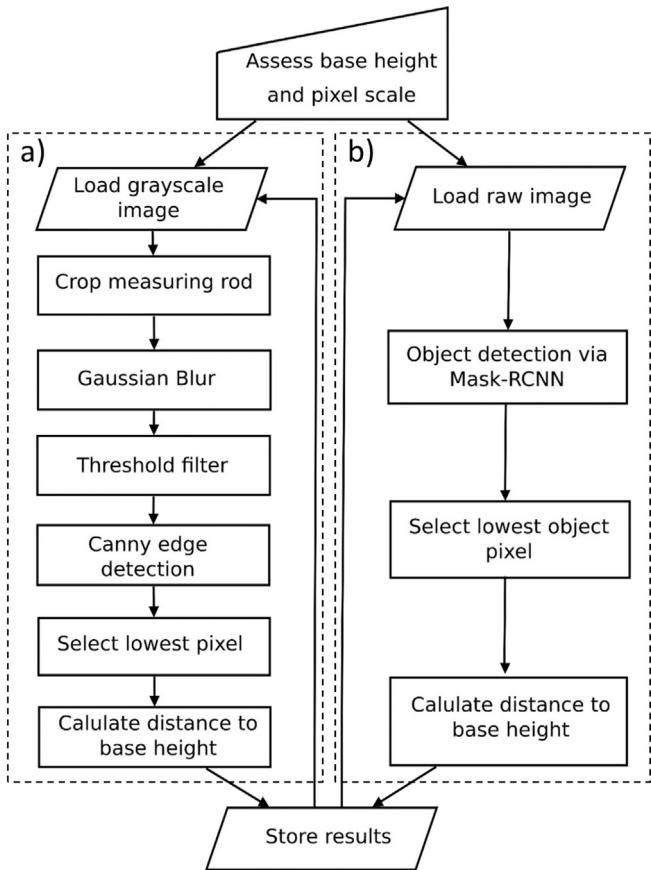


Fig. 2. Flow chart of the two automatic methods: a) IPM; b) MRN.

to measuring the vertical position of the lowest pixel of the measuring rod, the pixel scale in pixels per centimeter and the area of interest have to be assessed (see Fig. 3a). The parameters of the different steps have to be carefully adjusted to each location and season due to the

varying light situations. Once this parameter set is fixed for a study site, the automated measurement can be started: After loading the grayscale version of the image at hand, the area of interest is cropped (Fig. 3b). As a next step, a "Gaussian blur" is applied on the image in order to smear possible disturbances on the snow cover such as needles, leaves and contrast rich edges due to high snow surface roughness (Fig. 3c). In order to separate snow cover pixels from the measuring rod, a binary threshold is conducted (Fig. 3d). Finally, a canny edge filter is applied. This filter method locates the edges in the black and white image and therefore detects the outline of the measuring rod in front of the snow cover. From the resulting pixel matrix, we can extract the location of the lowest white pixel, which is to be found in the horizontal edge where the snow cover apparently meets the measuring rod. The vertical difference between the previously determined lowest pixel of the rod and this pixel multiplied with the pixel scale is the current snow depth.

2.3. Description of MRN

The machine learning method utilized in this investigation is a further developed framework based on a Convolutional Neural Network (CNN) (Krizhevsky et al., 2012) called "Mask R-CNN" (MRN) (He et al., 2017). It is a deep learning application that makes instance segmentation of objects in images possible, provided the neural network is trained properly. Instance segmentation combines the task of detecting the main objects in an image via bounding boxes (Girshick et al., 2014) with so called semantic segmentation, which aims to classify each pixel in a given set of categories (He et al., 2017). As a result, objects in an image are detected within their apparent boundaries, thus allowing for the determination of object sizes and distances in pictures. MRN is the latest offspring of the R-CNN family that was developed at "Facebook AI Research (FAIR)". It extends the Faster Region-based Convolutional Neural Network (Faster R-CNN) (Ren et al., 2015) by adding a procedure for predicting an object mask on pixel-level. The new branch works in parallel with the existing branches of the Faster R-CNN. It consists of a light weight fully connected network that is applied to each predicted Region of Interest (RoI) producing a pixel-based segmentation mask. This study investigates the usability of MRN for the determination of snow depths by localizing measuring

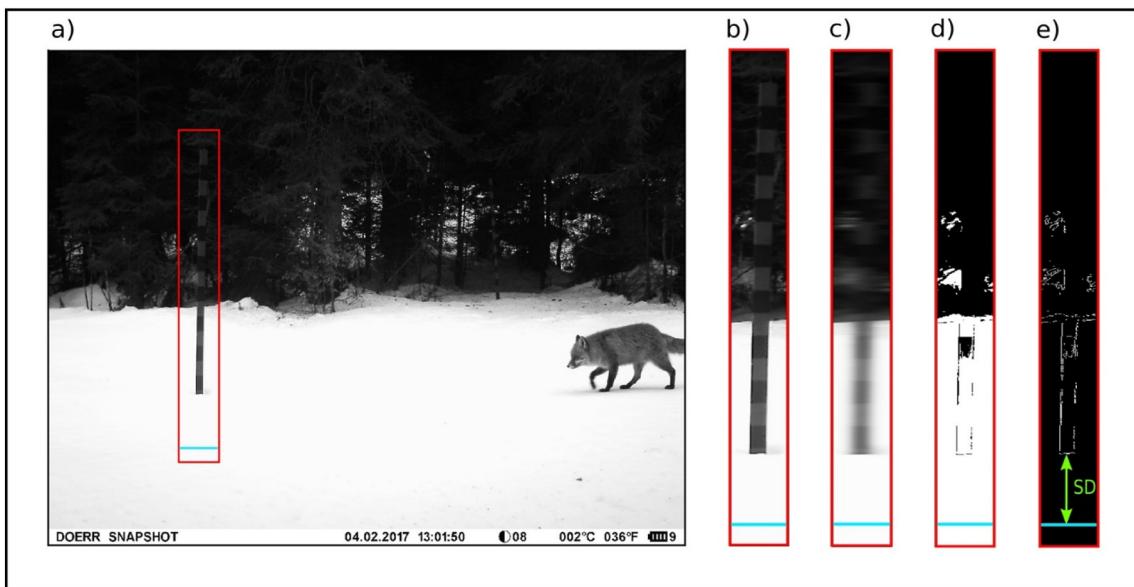


Fig. 3. IPM steps. The blue line indicates the bottom of the rod and the red box indicates the working area of the IPM. a) convert raw photo to grayscale; b) crop image area around the rod; c) apply Gaussian blur; d) application of the threshold filter; e) canny edge detection followed by calculation of the distance (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

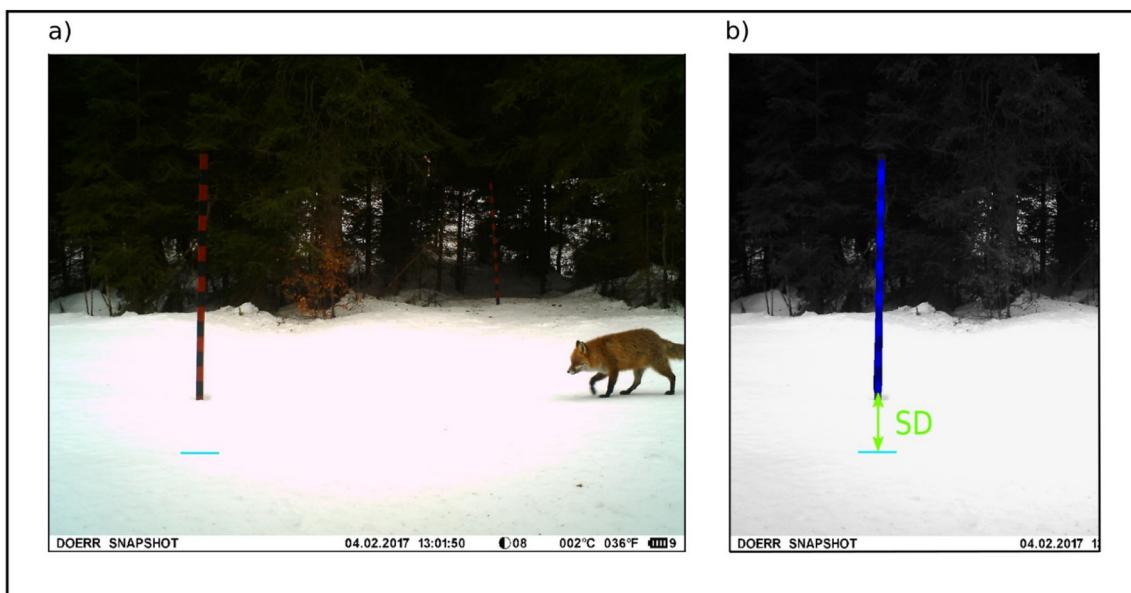


Fig. 4. MRN Process method. The blue line indicates the bottom of the rod. a) raw photo; b) calculation of snow depth value: the blue bar is the set of detected object pixels by MRN (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

rods (Fig. 4). Therefore, the neural network has been trained to detect the location of the visible (snow-free) part of a measuring rod in images as precisely as possible. As in the previously described IPM, the location of the lowest pixel of the rod is extracted from the segmentation mask and its vertical distance from the base of the rod. Finally, under consideration of the given pixel scale, the snow depth depicted in the image at hand can be determined. To train the MRN for segmenting instances of snow measuring rods in images, a collection of ~200 images of measurement rods have been utilized. The selected images cover day- and nighttime, multiple weather and light situations, different rod color combinations and various snow heights. For the training of the convolutional neural network utilized in this study, rods with the red/black and neon yellow/neon orange color combinations were used. A selection of the training set images used in this study is depicted in the supplementary material (Fig. S1). The necessary labeling process for the training data was performed using the “VGG Image Annotator” open-source software (Dutta et al., 2016).

2.4. Accuracy and error evaluation

The time series derived from manual measurements are considered as reliable references for evaluating the two automated processing methods. The accuracy of snow depth time series obtained from IPM and MRN methods are estimated quantitatively by calculating the root mean square error (RMSE), mean absolute error (MAE) and the standard deviation of the error (σ) against manual measurements (Harder et al., 2016; Ryan et al., 2008). In addition, the deviation between the automated and manual measurements is calculated for each time step, and the probability density is summarized in order to assess the error distribution. Any deviation larger than 20 cm is considered as large deviation. The ratio of the large deviation (RLD) is calculated for quantification analysis. For both study sites, assessment is carried out for 24-h, daytime only (9 a.m. to 4 p.m.), and nighttime only (4 p.m. to 9 a.m.) time series in order to compare the performance of the automated methods under different lighting situation. The 24-h dataset contains the data of the complete periods. Data of dusk and dawn periods are classified to nighttime dataset because they had similar bad light conditions. Daytime dataset excludes the night hours and the times of dusk and dawn.

3. Results

3.1. Performance of 24-h dataset

The 24-h snow depth time series are displayed in Figs. 5a and 6a. In both study sets, MRN performed very well during both day and night, as it was able to closely reproduce the behavior of the manual snow depth measurements with low errors and a very small ratio of large deviations indicated by Table 1. In contrast to MRN, IPM failed to continuously produce accurate snow depth values. IPM led to evidently high errors at nighttime period for both study sites. The number of large deviation was especially elevated with regard to the measurements performed at night. As a result, unsatisfactory error indices (RMSE, MAE, and σ) were obtained by IPM, which on average produced 3.7 and 8.2 times higher errors than MRN during the 24-h period at Rosseck and Talschluss respectively (Table 1). The occurrence frequency of large deviations was increased to a high level by IPM, with more than twenty and a hundred times higher RLD in comparison to the 24-h time series produced by MRN (Table 1). Systematic deviations in terms of numerous small fluctuations were observed in the datasets from both automated methods (Figs. 5 and 6). In general, the majority of these deviations are within 5 cm according to the deviation distributions shown in Fig. 7 (75.57% and 74.19% for IPM datasets, and 78.72% and 97.17% for MRN at Rosseck and Talschluss, respectively). MRN outperforms IPM in terms of error size, RLD and the amount of small deviations.

3.2. Performance of daytime dataset

IPM has presented bad performances to measure snow depth properly during night time (Figs. 5a, 6a and Table 1). In order to compare MRN and IPM thoroughly, in-depth evaluations were carried out by zooming in the performance of daytime periods when both methods worked properly. The daytime only time series were compared with the quantified support from error statistics in Figs. 5b and 6b (Table 1). During daytime, when the light situation is often satisfactory, both MRN and IPM produced snow depth data that fitted the manual measurements closely at both study sites. MRN slightly outperformed IPM in terms of errors and occurrence of large deviation: MRN produced zero large deviation during daytime, whereas IPM led to an RLD of 1.77%

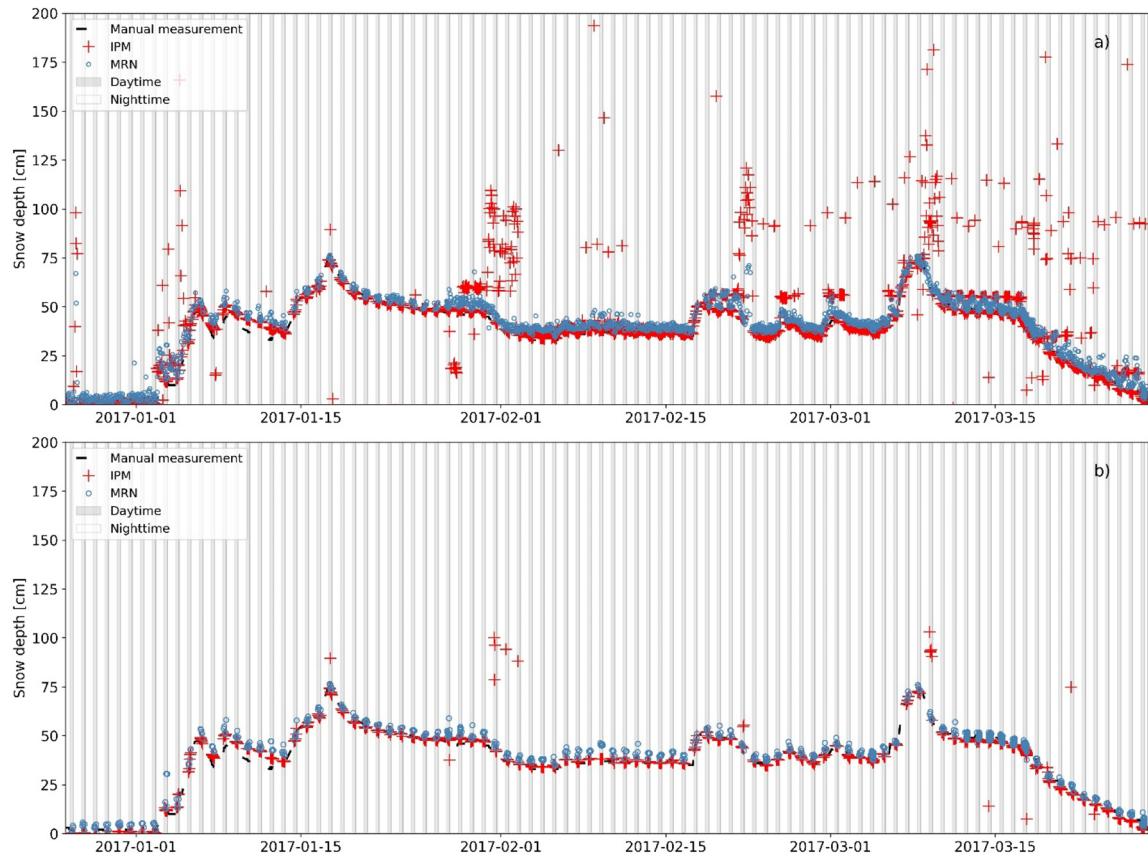


Fig. 5. Comparison of time series at the Rosseck station: a) 24-h, b) daytime.

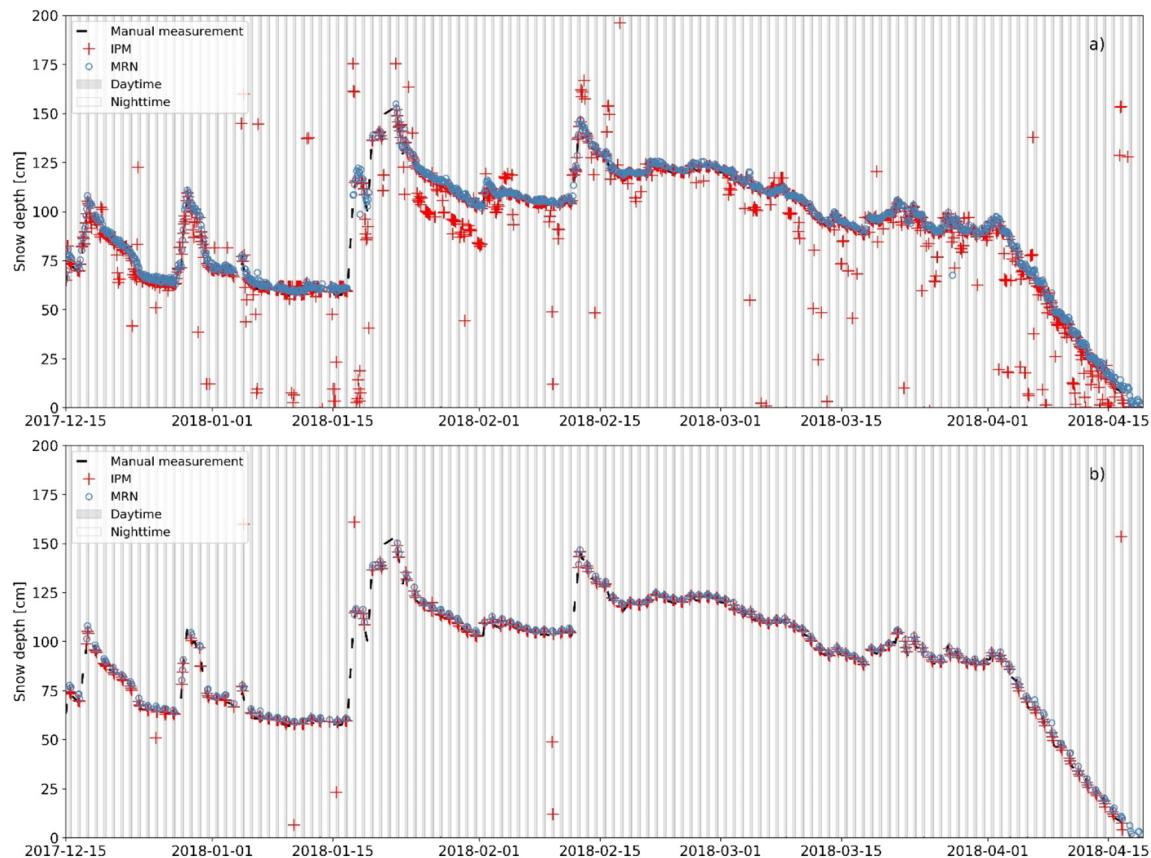


Fig. 6. Comparison of time series at the Talschluss station: a) 24-h, b) daytime.

Table 1
Evaluation index.

Dataset	Method	RMSE (cm)	MAE (cm)	σ (cm)	RLD
Rosseck (24-h)	Mask R-CNN	4.64	3.36	3.32	0.36%
	Image processing	17.79	6.89	17.15	8.78%
Rosseck (daytime)	Mask R-CNN	3.67	0.95	2.74	0
	Image processing	6.24	0.81	6.24	1.77%
Rosseck (nighttime)	Mask R-CNN	5.47	2.42	4.10	0.72%
	Image processing	21.60	5.98	20.44	12.38%
Talschluss (24-h)	Mask R-CNN	2.71	2.35	1.43	0.08%
	Image processing	19.93	7.80	19.73	8.71%
Talschluss (daytime)	Mask R-CNN	2.54	0.71	0.90	0
	Image processing	16.35	1.19	16.37	3.56%
Talschluss (nighttime)	Mask R-CNN	2.78	1.64	1.60	0.11%
	Image processing	21.30	6.61	20.93	10.95%

and 3.56% at Rosseck and Talschluss, respectively (Table 1). From the perspective of error metrics as shown in Table 1, MRN still showed small errors which were improved slightly compared to the 24-h performance. The improvements were evident for datasets produced by IPM at Rosseck: RMSE, MAE and σ were much smaller and remained in the same range as the results of MRN. For Talschluss as observed in Fig. 6b, the RMSE (16.35) and σ (16.37) are still large (Table 1) although the datasets of IPM look like a good match with the manual measurements. The reason is the dominant effect of the huge deviation of the countable large deviations (Fig. 6b), which is essentially ascribed to the fact that IPM failed to correctly extract the snow depth from blurred

photos (Fig. S2 in supplementary material) in bad weather conditions. Systematic small fluctuations were still noticeable for MRN datasets, especially at the Rosseck location, while the IPM time series were smooth. This is also reflected by the deviation distribution in Fig. 7: the percentages of the deviations within 5 cm are 85.62% and 93.57% for MRN and IPM at Rosseck. At Talschluss, 95% of the errors of both methods were small deviations <5 cm. In general, the two automated methods performed comparably well during daytime.

3.3. Cross validation of different study sites

MRN was trained to consider different measurement situations at different sites and different light situations. MRN offers a very high portability, since there are no further parameters that have to be adjusted to the trained neural network when a new measurement site is put into operation. As a consequence, the results of the MRN method remain unchanged during cross validation at both study sites (Fig. 8). This was not the case for IPM, which requires a carefully adjusted set of parameters for a specific measurement site. These parameters have to account for the differences in the lighting situation, the shape of the snow cover and the geometrical set-up of the measurement. This drawback severely reduces portability of this method. The cross validation of IPM parameters led to worse consistency with a high number of large deviation during day and night (Fig. 8). RLD was increased from 8.78% to 43.07% and from 8.71% to 48.59%, respectively, when Talschluss parameters were applied to Rosseck and conversely.

4. Discussion

MRN presented stable and good performances when it comes to extracting the snow depths from the image series of both study sites. Small errors were observed throughout all of the measurements, but

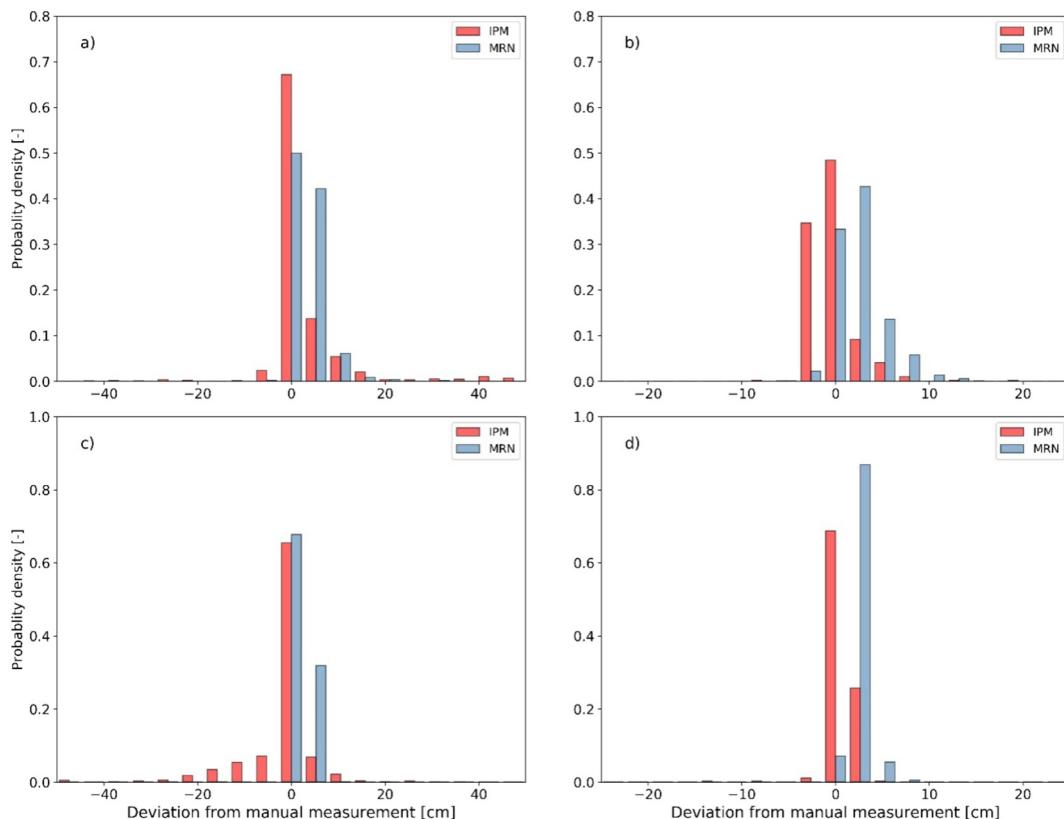


Fig. 7. Probability distribution of deviations from references of the two automated methods (“IPM/MRN” minus “manual measurements”): a) 24-h at Rosseck, b) daytime at Rosseck, c) 24-h at Talschluss, d) daytime at Talschluss.

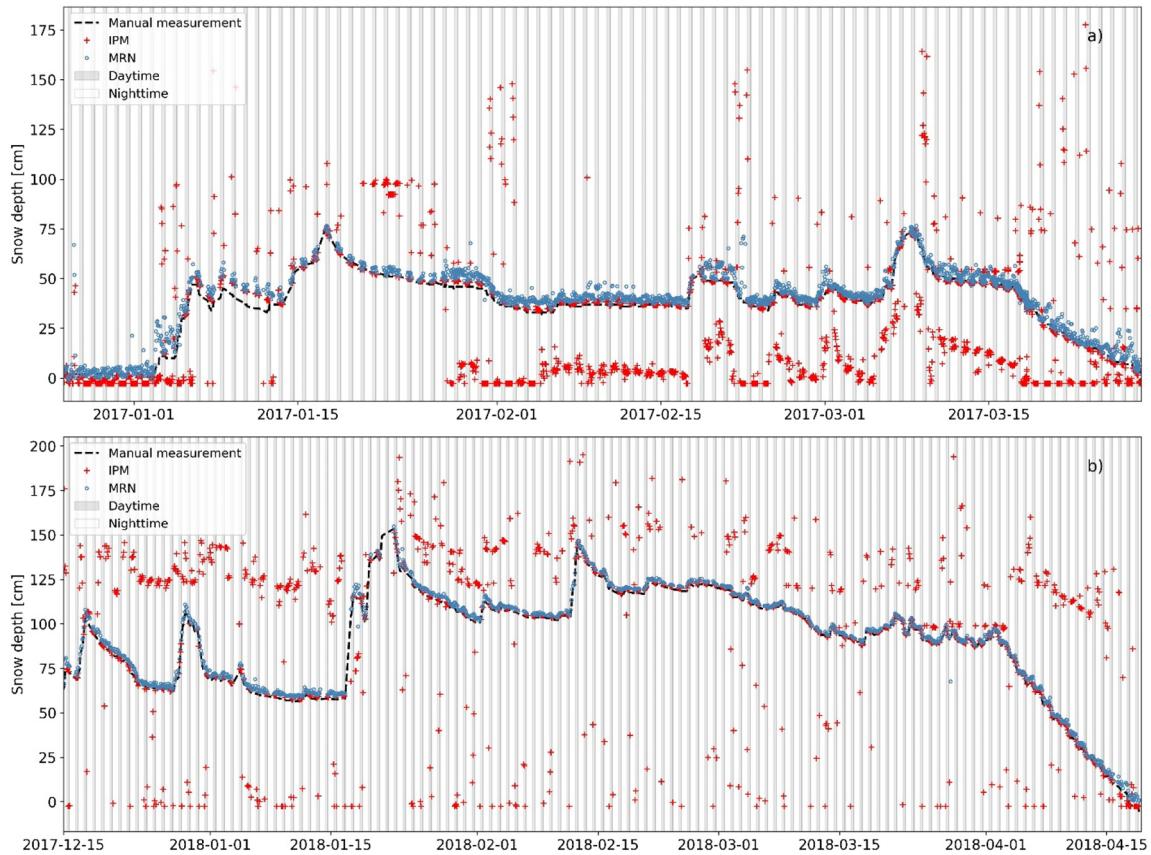


Fig. 8. Cross validation: a) Talschluss to Rosseck a) Rosseck to Talschluss.

MRN rarely (<0.5%) produced snow depth values that varied >20 cm from the manual measurements at both study sites. Systemic small deviations were persistently observed in terms of small fluctuations for snow depth time series of MRN. When zoomed into the processed photo of one "fluctuation" period (Fig. S3 in the supplementary material), the fluctuation is not only affected by the lighting situation in the image. In fact, the accuracy of MRN is limited by the joint impact of several factors: 1) contour detection was proven to be a challenging task for object segmentation using convolutional neural networks (Yang et al., 2016) and resulted in erroneous detection of the exact location of the apparent snowpack surface; 2) the quality of the optical system (lens) restricts photo quality, which increases the difficulty of detecting the edge accurately; 3) photo resolution is important; higher resolutions

generally lead to smaller deviations, which was observed at Talschluss (3264 pixel*2448 pixel) and Rosseck (2560 pixel *1920 pixel) respectively; 4) during the night, the photos turned out to be gray scale images due to bad lighting conditions. Red sections of the measuring bar appeared to be "Gray", which is close to the appearance of snow and distinct from black sections (Fig. S3 in the supplementary material). As a consequence, the red junction between the bar and the snow pack was not considered to be part of the measuring rod by the neural network, which led to an overestimation of snow depth. However, from a quantitative point of view, these deviations were mostly (>90% in average) <5 cm, which could be considered as reasonable and acceptable in snow dominated regions. Considering the measurement results obtained during daytime, IPM engendered measurements can be

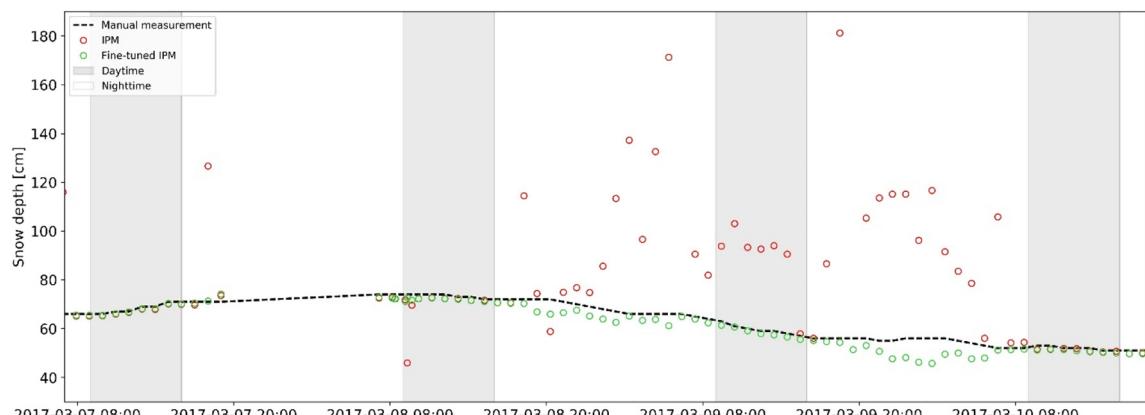


Fig. 9. Improvement of IPM after specific fine-tuning depending on time and weather.

considered as good as those of MRN, which are depicted in Figs. 5b and 6b (the curves are almost congruent) and confirmed by the corresponding statistics (Table 1).

However, IPM failed to produce continuously reliable 24-h snow depth data. Large deviations (bigger than 20 cm) occurred quite frequently for night datasets, since IPM could not properly deal with photos from night periods. The reason for this is the bad lighting situation during night, which makes it hard to find suitable filter adjustments that detect the apparent snowpack – measuring rod interface correctly. For a further investigation, specific parameter adjustments depending on time and weather were implemented for IPM for a short time period when a considerable amount of large deviation were produced at Rosseck (Fig. 9). As shown in Fig. 9, the improvements achieved by applying the fine-tuned IPM process are obvious. After applying the specific adjustments, large deviation (>20 cm deviation from reference) of the same time period were reduced from 37% to 0% (Fig. 9). Therefore, in order to further increase the performance of IPM, specific fine-tuning needs to be carried out for different lighting situations in different photos. For large photo series, this process is considered to be too time consuming.

In contrast, MRN is more user-friendly and less time-consuming. After being trained to work in different lighting situations, the neural network is able to detect the visible part of the measuring rod in new measurement situations, making MRN highly portable. MRN was able to produce stable snow depth data within an acceptable deviation interval compared to the manual measurements. It is widely suitable for various light situations and does not have to be modified. As a result, MRN led to acceptable snow depth values facing light variations from day to night, due to the topographic effects of different study sites, or because of extreme weather. Although the visual manual measurements are definitely the most accurate and reliable snow depth datasets, these are extremely time consuming (around 8.5 h per 1000 samples). This situation makes manual measurements not suitable for large datasets or urgent need of data. Conversely, MRN is highly efficient (about 3 h per 1000 samples on 12 Intel (R) Xeon (R) CPU E5-1650 v3 @ 3.50GHz cores) and produces good data mainly within a 5 cm deviation from the manual measurements. Utilizing a fast GPU will substantially decrease the necessary computation time. Therefore, MRN is an appropriate alternative for manual measurements when large data sets have to be considered or when evaluation needs to be rapid.

Since IPM requires specific parameterization fitting specific lighting conditions, its use is not recommended for acquiring data in complex natural conditions where lighting variation is large. We don't exclude the possibility of improving the IPM method with an automated parameter adaption algorithm to automatically cope with the lighting variation. This is a standalone research topic and therefore beyond the scope of this work.

Besides, both methods have systematic errors which lead to systematic underestimation of IPM and overestimation of MRN (Figs. 5–7). Due to disturbances (needles, leaves, light disturbances) on the snow cover, the IPM method often falsely detects lower edges applying the "canny edges" filter and hence is underestimating the snow depth. Whereas the MRN method overestimated the snow depth because of the incomplete instance segmentation of the measuring rod (Fig. S3 in the supplementary material). Especially in the boundary areas of the measuring rods, the MRN tended to fail in the recognition of pixels as part of the rod. Therefore, the visible part of the rod appeared to be shorter. This behavior led to a tendency to overestimate the snow depth by MRN.

5. Conclusion

Efficient and reliable snow depth measurement is important for water resource management in snow dominated regions. With the rapid development of machine learning methods, convolutional neural network frameworks such as Mask R-CNN have become an available tool for collecting detailed information from photos that could greatly

accelerate the working efficiency of laborious and time-consuming manual work. In this study, the MRN workflow that uses the Mask R-CNN framework has been proven to be feasible and efficient to produce snow depth datasets that are close to manual references based on hourly photo series of time-lapse cameras. In a comparison of these with manual measurements, MRN showed great competence in continuously obtaining accurate enough snow depth data under various lighting conditions due to day-night alterations, topography shading, or different weather situations. Its training process is simple and user-friendly. Therefore, MRN is a highly flexible framework suitable for instance segmentation at a variety of study sites. Although the IPM method could also produce some comparably good snow depth data, it demands even more time than the manual measurement because IPM needs very specific parameterization in order to deal with difficult light conditions. MRN is therefore suggested to be useful and highly efficient for the collection of long-term snow depth data from image series recorded by time-lapse cameras. In this work, we trained Mask R-CNN with many types of measurement situations obtained from the measurement network in the Dreisäuerbach catchment. As an outlook, it would be helpful to evaluate the performance curve of MRN against the amount of considered training samples. This could be an interesting and standalone topic for future investigations, but is out of the scope of this work. Convolutional neural networks such as Mask R-CNN are powerful tools for acquiring information from various images in the field of environmental research (Abolt et al., 2019; Zhang et al., 2018). There was a very recent attempt to make use of deep neural networks for flood area detection (Moy de Vitry et al., 2019). In addition, a conceivable future application of MRN could be the monitoring of water levels in urban areas during flooding. Measuring rods could be painted on walls and recorded in image series from CCTV camera systems. The trained Mask R-CNN network of this study can easily be extended to detect these painted rods. As a result, a real-time and continuous urban water level map could be provided to the water authorities, government and scientists to grasp flood situations and implement a fast flood control response.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was part of the ProMoS (Process based modeling of snow cover) Project, funded by the Bavarian State Ministry of the Environment and Consumer Protection. We thank Dr. Jakob Garvelmann (formerly Karlsruhe Institute of Technology Campus Alpin, Garmisch-Partenkirchen) for providing the image series of the measurements at the "Talschluss" location. The authors would like to thank the Associate Editor Prof. Dr. Ralf Ludwig and the anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2019.134213>.

References

- Abolt, C.J., Young, M.H., Atchley, A.L., Wilson, C.J., 2019. Brief communication: rapid machine-learning-based extraction and measurement of ice wedge polygons in high-resolution digital elevation models. *Cryosphere* 13 (1), 237–245. <https://doi.org/10.5194/tc-13-237-2019>.
- Bell, V.A., Kay, A.L., Davies, H.N., Jones, R.G., 2016. An assessment of the possible impacts of climate change on snow and peak river flows across Britain. *Clim. Chang.* 136 (3), 539–553. <https://doi.org/10.1007/s10584-016-1637-x>.

- Boniface, K., Braun, J.J., McCreight, J.L., Nievinski, F.G., 2015. Comparison of Snow Data Assimilation System with GPS reflectometry snow depth in the Western United States. *Hydrol. Process.* 29 (10), 2425–2437. <https://doi.org/10.1002/hyp.10346>.
- Bradski, G., 2000. The opencv library. *Dr. Dobb's J. Softw. Tools* 25, 120–125.
- Chang, A.T.C., et al., 2005. Analysis of ground-measured and passive-microwave-derived snow depth variations in midwinter across the Northern Great Plains. *J. Hydrometeorol.* 6 (1), 20–33. <https://doi.org/10.1175/jhm-405.1>.
- Deems, J.S., Fassnacht, S.R., Elder, K.J., 2006. Fractal distribution of snow depth from Lidar data. *J. Hydrometeorol.* 7 (2), 285–297. <https://doi.org/10.1175/jhm487.1>.
- Deems, J.S., Painter, T.H., Finnegan, D.C., 2013. Lidar measurement of snow depth: a review. *J. Glaciol.* 59 (215), 467–479. <https://doi.org/10.3189/2013JoG12J154>.
- Dong, C., Menzel, L., 2017. Snow process monitoring in montane forests with time-lapse photography. *Hydrol. Process.* 31 (16), 2872–2886. <https://doi.org/10.1002/hyp.11229>.
- Dutta, A., Gupta, A., Zissermann, A., 2016. VGG Image Annotator (VIA). <http://www.robots.ox.ac.uk/~vgg/software/via>.
- Farinotti, D., Magnusson, J., Huss, M., Bauder, A., 2010. Snow accumulation distribution inferred from time-lapse photography and simple modelling. *Hydrol. Process.* 24 (15), 2087–2097. <https://doi.org/10.1002/hyp.7629>.
- Garvelmann, J., Pohl, S., Weiler, M., 2013. From observation to the quantification of snow processes with a time-lapse camera network. *Hydrol. Earth Syst. Sci.* 17 (4), 1415–1429. <https://doi.org/10.5194/hess-17-1415-2013>.
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 580–587.
- Gutmann, E.D., Larson, K.M., Williams, M.W., Nievinski, F.G., Zavorotny, V., 2012. Snow measurement by GPS interferometric reflectometry: an evaluation at Niwot Ridge, Colorado. *Hydrol. Process.* 26 (19), 2951–2961. <https://doi.org/10.1002/hyp.8329>.
- Harder, P., Schirmer, M., Pomeroy, J., Helgason, W., 2016. Accuracy of snow depth estimation in mountain and prairie environments by an unmanned aerial vehicle. *Cryosphere* 10 (6), 2559–2571. <https://doi.org/10.5194/tc-10-2559-2016>.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. 2017. *IEEE Int. Conf. Comput. Vis.*, 2980–2988 <https://doi.org/10.1109/ICCV.2017.322>.
- Horton, P., Schaeffer, B., Mezghani, A., Hingray, B., Musy, A., 2006. Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrol. Process.* 20 (10), 2091–2109. <https://doi.org/10.1002/hyp.6197>.
- Kinar, N.J., Pomeroy, J.W., 2015. Measurement of the physical properties of the snowpack. *Rev. Geophys.* 53 (2), 481–544. <https://doi.org/10.1002/2015RG000481>.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 1097–1105.
- Kwok, R., Kacimi, S., 2018. Three years of sea ice freeboard, snow depth, and ice thickness of the Weddell Sea from Operation IceBridge and CryoSat-2. *Cryosphere* 12 (8), 2789–2801. <https://doi.org/10.5194/tc-12-2789-2018>.
- Lawrence, D., Paquet, E., Gailhard, J., Fleig, A.K., 2014. Stochastic semi-continuous simulation for extreme flood estimation in catchments with combined rainfall-snowmelt flood regimes. *Nat. Hazards Earth Syst. Sci.* 14 (5), 1283–1298. <https://doi.org/10.5194/nhess-14-1283-2014>.
- Leppänen, L., Kontu, A., Hannula, H.R., Sjöblom, H., Pulliainen, J., 2016. Sodankylä manual snow survey program. *Geosci. Instrum. Method. Data Syst.* 5 (1), 163–179. <https://doi.org/10.5194/gi-5-163-2016>.
- McGrath, D., et al., 2015. End-of-winter snow depth variability on glaciers in Alaska. *J. Geophys. Res. Earth Surf.* 120 (8), 1530–1550. <https://doi.org/10.1002/2015JF003539>.
- Moy de Vitry, M., Kramer, S., Wegner, J.D., Leitão, J.P., 2019. Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network. *Hydrol. Earth Syst. Sci. Discuss.* (20), 19: 1–21 <https://doi.org/10.5194/hess-2018-570>.
- Parajka, J., Haas, P., Kirnbauer, R., Jansa, J., Blöschl, G., 2012. Potential of time-lapse photography of snow for hydrological purposes at the small catchment scale. *Hydrol. Process.* 26 (22), 3327–3337. <https://doi.org/10.1002/hyp.8389>.
- Pirazzini, R., et al., 2018. European in-situ snow measurements: practices and purposes. *Sensors* 18 (7), 2016.
- Prokop, A., 2008. Assessing the applicability of terrestrial laser scanning for spatial snow depth measurements. *Cold Reg. Sci. Technol.* 54 (3), 155–163. <https://doi.org/10.1016/j.coldregions.2008.07.002>.
- Prokop, A., Schirmer, M., Rub, M., Lehning, M., Stocker, M., 2008. A comparison of measurement methods: terrestrial laser scanning, tachymetry and snow probing for the determination of the spatial snow-depth distribution on slopes. *Ann. Glaciol.* 49, 210–216. <https://doi.org/10.3189/172756408787814726>.
- Raleigh, M.S., Lundquist, J.D., Clark, M.P., 2015. Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.* 19 (7), 3153–3179. <https://doi.org/10.5194/hess-19-3153-2015>.
- Rasmussen, R., et al., 2012. How well are we measuring snow: the NOAA/FAA/NCAR winter precipitation test bed. *Bull. Am. Meteorol. Soc.* 93 (6), 811–829. <https://doi.org/10.1175/bams-d-11-00052.1>.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, pp. 91–99.
- Revuelto, J., Jonas, T., López-Moreno, J.I., 2016. Backward snow depth reconstruction at high spatial resolution based on time-lapse photography. *Hydrol. Process.* 30 (17), 2976–2990. <https://doi.org/10.1002/hyp.10823>.
- Ryan, W.A., Doesken, N.J., Fassnacht, S.R., 2008. Evaluation of ultrasonic snow depth sensors for U.S. snow measurements. *J. Atmos. Ocean. Technol.* 25 (5), 667–684. <https://doi.org/10.1175/2007jtecha947.1>.
- Schattan, P., et al., 2017. Continuous monitoring of snowpack dynamics in alpine terrain by aboveground neutron sensing. *Water Resour. Res.* 53 (5), 3615–3634. <https://doi.org/10.1002/2016WR020234>.
- Shapiro, L.G., Stockman, G., 2001. *Computer Vision*. Prentice Hall PTR 608 pp.
- Shinohara, Y., Kumagai, T., Otsuki, K., Kume, A., Wada, N., 2009. Impact of climate change on runoff from a mid-latitude mountainous catchment in central Japan. *Hydrol. Process.* 23 (10), 1418–1429. <https://doi.org/10.1002/hyp.7264>.
- Shrestha, R., Takara, K., Tachikawa, Y., Jha, R.N., 2004. Water resources assessment in a poorly gauged mountainous catchment using a geographical information system and remote sensing. *Hydrol. Process.* 18 (16), 3061–3079. <https://doi.org/10.1002/hyp.5749>.
- Sturm, M., Holmgren, J., 2018. An automatic snow depth probe for field validation campaigns. *Water Resour. Res.*, 54(0). DOI:<https://doi.org/10.1029/2018WR023559>.
- Szczypta, C., et al., 2015. Impact of climate and land cover changes on snow cover in a small Pyrenean catchment. *J. Hydrol.* 521, 84–99. <https://doi.org/10.1016/j.jhydrol.2014.11.060>.
- Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018a. Calibration of snow parameters in SWAT: comparison of three approaches in the Upper Adige River basin (Italy). *Hydrol. Sci. J.*, 1–22 <https://doi.org/10.1080/02626667.2018.1439172>.
- Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018b. A multi-objective approach to improve SWAT model calibration in alpine catchments. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2018.02.055>.
- Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D., Wong, W.K., 2016. Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway. *J. Hydrol.* 538, 33–48. <https://doi.org/10.1016/j.jhydrol.2016.03.066>.
- Warscher, M., et al., 2013. Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. *Water Resour. Res.* 49 (5), 2619–2637. <https://doi.org/10.1002/wrr.20219>.
- Xuejin, T., et al., 2018. Spatiotemporal changes in snow cover over China during 1960–2013. *Atmos. Res.* <https://doi.org/10.1016/j.atmosres.2018.11.018>.
- Yang, J., Price, B., Cohen, S., Lee, H., Yang, M.-H., 2016. Object contour detection with a fully convolutional encoder-decoder network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 193–202.
- Zhang, W., Witharana, C., Liljedahl, A., Kanevskiy, M., 2018. Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery. *Remote Sens.* 10 (9), 1487.
- Zierl, B., Bugmann, H., 2005. Global change impacts on hydrological processes in Alpine catchments. *Water Resour. Res.* 41 (2), 1487. <https://doi.org/10.1029/2004wr003447>.