LAKE ICE MONITORING WITH WEBCAMS

Muyan Xiao, Mathias Rothermel[*], Manu Tom, Silvano Galliani, Emmanuel Baltsavias, Konrad Schindler

Photogrammetry and Remote Sensing, ETH Zürich, firstname.lastname@geod.baug.ethz.ch

Commission II, WG II/6

KEY WORDS: Climate Monitoring, Lake Ice Monitoring, Webcams, Semantic Segmentation, Convolutional Neural Networks

ABSTRACT:

Continuous monitoring of climate indicators is important for understanding the dynamics and trends of the climate system. Lake ice has been identified as one such indicator, and has been included in the list of Essential Climate Variables (ECVs). Currently there are two main ways to survey lake ice cover and its change over time, in-situ measurements and satellite remote sensing. The challenge with both of them is to ensure sufficient spatial and temporal resolution. Here, we investigate the possibility to monitor lake ice with video streams acquired by publicly available webcams. Main advantages of webcams are their high temporal frequency and dense spatial sampling. By contrast, they have low spectral resolution and limited image quality. Moreover, the uncontrolled radiometry and low, oblique viewpoints result in heavily varying appearance of water, ice and snow. We present a workflow for pixel-wise semantic segmentation of images into these classes, based on state-of-the-art encoder-decoder Convolutional Neural Networks (CNNs). The proposed segmentation pipeline is evaluated on two sequences featuring different ground sampling distances. The experiment suggests that (networks of) webcams have great potential for lake ice monitoring. The overall per-pixel accuracies for both tested data sets exceed 95%. Furthermore, per-image discrimination between ice-on and ice-off conditions, derived by accumulating per-pixel results, is 100% correct for our test data, making it possible to precisely recover freezing and thawing dates.

1. INTRODUCTION

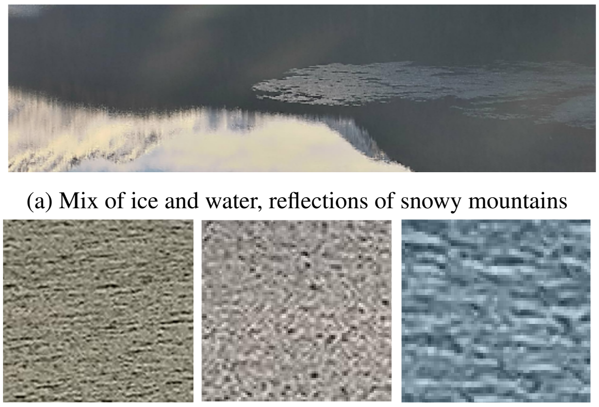
Climate change and global warming significantly impact the environment and human livelihoods. Hence, there is a need to monitor and understand the climate system and its important parameters. While there is not yet an exhaustive list of parameters that must be recorded to characterize the global climate, lake ice is known to closely follow the temporally integrated air temperature and has long been recognized as an important indicator of climate change (Robertson et al., 1992, Latifovic and Pouliot, 2007, Brown and Duguay, 2010). To support climate research, the World Meteorological Organization and other related organizations have established a database termed the "Global Climate Observing System" (GCOS), with the aim of providing world-wide records of the most significant physical, biological and chemical variables, the so-called *Essential Climate Variables* (ECVs). Lake ice cover is one such variable within the category "lakes", with the key measurements being the spatial extend of ice coverage along with its temporal changes, i.e., freezing and thawing dates. The work described in this paper forms part of a project to identify suitable sensors and processing methods for automatic ice monitoring on Swiss lakes, initiated by the Federal Office of of Meteorology and Climatology (MeteoSwiss).

Directly measuring temperature close to the water surface is perhaps the most intuitive way to survey lake ice. However, measurements of sensors placed very near to the water surface are heavily biased by the temperature of the ambient air. Probes placed below water level do not allow for a reliable retrieval of ice coverage at the surface. Another challenge is the installation and maintenance of a dense sensor network, which is costly and in many cases impractical due to the harsh environment and conflicts with the use of water bodies, e.g., for shipping. Lake ice monitoring by satellite remote sensing is based on either optical or microwave imagery. For an overview of sensors and methods used to survey river and inland ice, refer to (Duguay et al., 2015). The main disadvantage of remote sensing is its limited spatial and temporal resolution. In particular, there is a trade-off between high spatial resolution (only possible with small sensor footprints) and high temporal resolution (requiring frequent revisits). For optical sensors, temporal resolution is further impaired by cloud coverage. While some promising work exists, e.g., (Sütterlin et al., 2017, Tom et al., 2017), lake ice monitoring with satellite data struggles to fulfill even the current ECV specifications, which demand daily observations at 300 meter GSD.

On the contrary, ground-based webcams provide excellent spatial and temporal resolution, and are cheap and easy to install. Moreover, a rather dense network of cameras already exists, many of which allow access to the data streams via public web services. Note that in some parts of the world (including Switzerland) this is particularly true for lakes, due to their value for recreation, tourism, energy production, etc. Potential drawbacks of webcams are the incomplete coverage of many lake surfaces, as well as temporal data gaps due to dense fog or heavy rain and snowfall. For our test site, the moderate-sized lake of St. Moritz, publicly available webcams cover the entire water surface. We note that in mountain areas (like Switzerland), many lakes are surrounded by steep terrain, making it easy to install cameras at appropriate, elevated viewpoints with wide field-of-view, so as to improve coverage.

In this article we investigate the potential of RGB webcam images to predict accurate, per-pixel lake ice coverage. Technically, this amounts to a semantic segmentation of the image into the classes *water*, *ice*, *snow* and *clutter*, which we implement with a state-of-the-art deep convolutional neural network (CNN). The *snow* class is necessary to cover the case where snow covers the ice layer, whereas clutter accounts for objects other than the three target classes that may temporally appear on a lake. The key

[*] Corresponding author



(a) Mix of ice and water, reflections of snowy mountains

(b) Water texture (c) Water texture (d) Ice texture

Figure 1. Examples of lake textures observed with webcams.

challenge when working with cheap webcams in outdoor conditions is the data quality, as highlighted in Figure 1. The low viewpoints lead to large variations in perspective scale, the uncontrolled lighting and weather conditions cause specular reflections, moving shadows and strong appearance differences within the same class, while the image quality is also limited (low signal-to-noise ratios, compression artifacts). In some cases, even manual classification is difficult and only possible by exploiting temporal cues. Despite these circumstances, we find that excellent segmentation results can be obtained with modern CNNs. While the core of our system is yet another variant of the recently successful *DenseNet/Tiramisu* architecture; there is, to the best of our knowledge, no published work regarding lake ice monitoring with webcams or other terrestrial photographs. Looking beyond lake ice and at environmental monitoring in general, we find that webcams are still an under-utilized resource, and that deep learning could also benefit many other environmental applications. We thus hope our study will trigger further work in this direction.

2. RELATED WORK

2.1 Terrestial and Webcam Data for Environmental Monitoring

Many environmental monitoring applications use image sequences captured with ground-based cameras, including vegetation phenology, fog monitoring, cloud tracking, rain- and snowfall assessment and estimation of population size, to name a few. For an excellent overview see (Bradley and Clarke, 2011). As pointed out by (Jacobs et al., 2009), dense webcam networks constitute an interesting alternative to remote sensing data to retrieve environmental information. Besides presenting two webcam-based algorithms to estimate weather signals and temporal properties of spring leaf growth, the authors maintain the *Archive of Many Outdoor Scenes* (AMOS) (Jacobs et al., 2007), which collects imagery from nearly 30000 webcams world-wide. (Richardson, 2015) present a continental-scale dataset consisting of 200 cameras, specifically tailored for research in vegetation phenology.

In the following we concentrate on methods for pixel-wise classification in the context of environmental applications. An algorithm for monitoring canopy phenology from webcam imagery was presented in (Richardson et al., 2007), which fits a sigmoid model to entities computed from the raw RGB information. (Bothmann et al., 2017) propose a semi-supervised and an unsupervised approach to identify regions in webcam streams that depict vegetation. Phenology of the vegetation is then assessed by tracking temporal changes in the green channel. In the domain of snow monitoring, (Salvatori et al., 2011, Arslan et al., 2017) present methods to estimate snow coverage in image sequences. Pixel-wise classification is done by thresholding intensity with a threshold value derived from the histogram of the blue channel. (Rüfenacht et al., 2014) fit a Gaussian Mixture Model to classify snow pixels, and enforce spatial and temporal consistency of segmentations via a Markov Random Field. (Fedorov et al., 2016) train binary snow-on/snow-off classification with a Random Forest and Support Vector Machines. Using a 33-dimensional feature vector, their supervised methods outperform thresholding as in (Salvatori et al., 2011).

Perhaps the closest work to ours is (Bogdanov et al., 2005), where a shallow neural network is trained to classify feature vectors extracted from SAR and optical satellite imagery as well as terrestrial photographs. The network predicts 6 classes of sea ice with an overall accuracy of approximately 91%. To the best of our knowledge, no work exists about lake ice detection based on terrestrial images.

2.2 CNNs for Semantic Segmentation

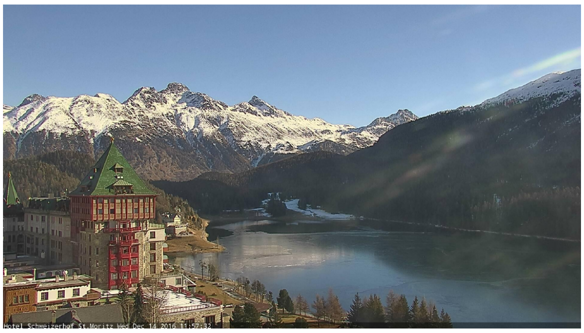
The rise of deep neural networks for image processing has recently also boosted semantic image segmantation. Based on the seminal *Fully Connected Network* of (Long et al., 2015), many state-of-the-art segmentation networks follow the encoder-decoder architecture. The encoder is typically derived from some high-performance classification network consisting of a series of convolution (followed by non-linear transformations) and downsampling layers, for instance (He et al., 2015, Huang et al., 2016, Xie et al., 2016). The subsequent decoder uses transposed convolutions to perform upsampling, normally either reusing higher-resolution feature maps (Long et al., 2015, Ronneberger et al., 2015, Jégou et al., 2016) or storing the pooling patterns of the encoder (Badrinarayanan et al., 2015). In this way, the high-frequency details of the input image can be recovered. The present work builds on the *Tiramisu* network proposed in (Jégou et al., 2016), which we will review in more detail in section 3.2.

3. METHODOLOGY

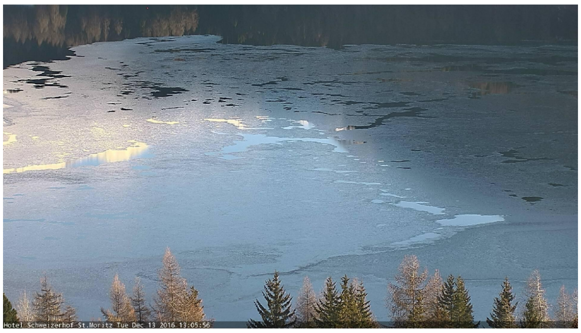
3.1 Data Collection and Preprocessing

The data used in this work consists of image streams from two webcams, which we have automatically downloaded from the internet. Both cameras capture lake St. Moritz, see Figures 2a and 2b. Images were collected from December 2016 until June 2017. The lake was frozen for a period of approximately four months, starting mid-December. The major difference between the two streams are image scale: one camera (Cam0) captured images with larger GSD whereas the other one (Cam1) records at higher resolution. Both cameras record at a frequency of one image per hour. The cameras are stationary and stable with respect to wind, such that the maximal movements observed in the data are around 1 pixel. We manually removed images affected by heavy snow fall, fog and bad illumination conditions (early morning, late evening). Methods for automatic detection and elimination of such images have been proposed, e.g. (Fedorov et al., 2016), but are not in the scope of this work.

Ground truth label maps were produced by manually delineating and labeling polygons in the images, with labels *water*, *ice*, *snow* and *clutter*. Among these,*water*, *ice* and *snow* are the sought



(a) Cam0 - large GSD stream



(b) Cam1 - small GSD stream

Figure 2. Example images of the two webcam streams.

attributes of the application, the *clutter* class was introduced to mark objects other than water that are sometimes found on the lake, such as boats, or tents which are built up on lake St.Moritz when hosting horse racing events. For the manual labeling task we used the browser-based tool of (Dutta et al., 2016). The specified polygons were then converted to raster label maps with a standard point-in-polygon algorithm. Overall, 820 images for Cam0 and 927 images for Cam1 were labeled.

3.2 Semantic Segmentation

Our segmentation network is based on the *One Hundred Layer Tiramisu* architecture of (Jégou et al., 2016). The network features a classical encoder-decoder architecture, see Figure 3(a). The encoder is based on the classification architecture *DenseNet*, a sequence of so-called dense blocks (DB), see Figure 3(b). A dense block contains several layers. Each layer transforms its input by batch normalization (Ioffe and Szegedy, 2015), ReLU rectification (Glorot et al., 2011) and convolution. The depth of the convolution layer is called *growth rate*. The distinguishing characteristic of a dense block is that the result of the transformation is *concatenated* with the input to form the output that is passed to the next layer, thus propagating lower-level representations up the network. In much the same way, the output of a complete dense block is concatenated with its input and passed through a transition-down (TD) block to reduce the resolution. TD blocks are composed of batch normalization, ReLU, 3×3 convolution and average-pooling. To make the model more compact, the 3×3 convolution reduces the depth of the feature maps by a fixed *compression rate*. The result is then fed into the next dense block. The input feature maps of each transition-down block are also passed to the decoder stage with the appropriate resolution, to better recover fine details during up-sampling. The decoder is a sequence of dense blocks and transition-up (TU) blocks. Note that in contrast to the encoder, dense blocks pass only the transformed feature maps, but not their inputs, to the next stage, to control model complexity. Transition-up blocks are composed of transposed convolutions with stride 2, which perform the actual up-sampling. Output feature maps from the last dense block are subject to a final reduction in depth, followed by a softmax layer to obtain probabilities for each class at each pixel. The connection between the encoder and the decoder part is one more dense block (*bottleneck*), which has the lowest spatial resolution and at the same time the highest layer depth. It can be interpreted as a sort of abstract “internal representation” shared by the input data and the segmentation map.

In practice, the input dimensions are limited by the available GPU memory. To process complete images, we cut them into 224×224 pixel tiles with 50% overlap along the row and column direction, such that each pixel is contained in 4 tiles. Each tile is processed separately, then the four predicted probabilities $p^c_{i=0,1,2,3}(\mathbf{x})$ for class c are averaged at every pixel $\mathbf{x}$, to obtain $p^c(\mathbf{x}) = \sum_i p^c_i(\mathbf{x})/4$. The final class is then the one with highest probability (winner-takes-it-all).

4. EXPERIMENTS

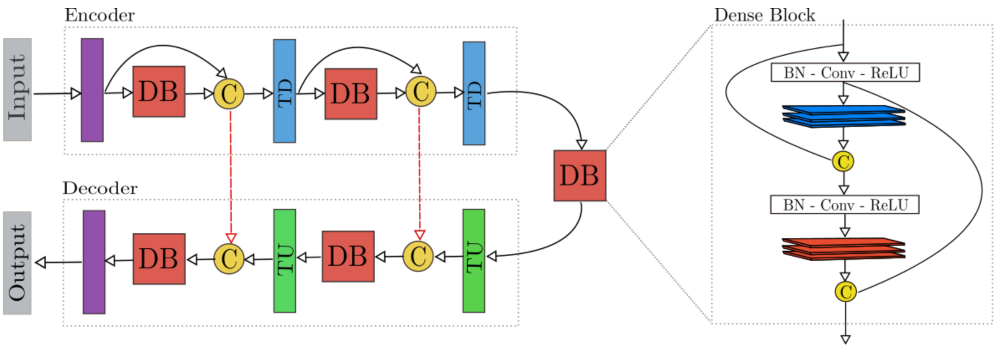
4.1 Training Details

Training and test sets are generated by randomly selecting 75% of all images for training and the remaining 25% for testing. All images are then tiled into 224×224 patches as described in section 3.2. The set of training patches is further subdivided (randomly) into a training part (80% of training data, respectively 60% of all data) and a validation part (20%, respectively 15%). All patches are normalized by subtracting the mean intensity. Class frequencies are balanced in the cross-entropy loss function by reweighting with the (relative) frequencies in the training set. The same network architecture is used for both cameras. It features three dense blocks in the encoder (with 4,7 and 12 layers), and three dense blocks in the decoder (with 12,7, and 4 layers). The bottleneck which connects encoder and decoder has 15 layers. The *growth rate* is 12. Learning is done with the Nestorov-Adam optimizer (Sutskever et al., 2013). The network is regularized with L2-regularization and dropout (Srivastava et al., 2014) with a rate of 50%. We found empirically that high *compression rates* of 0.25 to 0.33 were important to ensure good convergence. The network was implemented using *Keras*, with *Tensorflow* as backend. All experiments were run on a *Nvidia Titan X* graphics card.

4.2 Quantitative Results: Semantic Segmentation

We train separate networks (i.e., same architecture, but individual network weights) for the two datasets, so as to adapt the network weights to the specific camera and viewpoint. After training, the network is applied to all test patches of the respective dataset, and the patch-wise predictions are assembled to complete per-image segmentation maps with the consensus mechanism explained in Section 3.2. A background mask is applied to the images so that only pixels which correspond to the water body are evaluated. The resulting pixel-wise class maps per full camera image are the final predictions that we compare to ground truth. The confusion matrices for the two datasets are displayed in Tables 1a and 1b. Entries are absolute pixel counts across the entire test set, in units of 1 million pixels. Furthermore, we also display precision and recall for each class, as well as the overall accuracy.

The segmentation results are promising, reaching overall accuracies of 95.3% for the Cam0 sequence and 95.7% for the Cam1 sequence. For both datasets, semantic segmentation of *water* works



(a) (b)

Figure 3. (a): Schematic illustration of the segmentation framework. The encoder down-samples the image, and thereby increases the field of view. It consists of a sequence of dense blocks (DB) and transition down (TD) blocks. The decoder performs up-sampling of feature maps with a sequence of dense blocks (DB) and transition-up (TU) blocks. To recover high-resolution detail, skip connections pass information from intermediate encoder stages to the corresponding decoder stages. (b): Internal structure of a dense block with two layers and *growth rate* 3. For more details, see Section 3.2.

Predicted True	water	ice	snow	clutter	Recall
water	**28.57**	0.29	0.31	0.00	98.0%
ice	0.40	**4.56**	0.20	0.00	88.3%
snow	0.51	0.20	**8.29**	0.07	91.4%
clutter	0.00	0.00	0.08	**0.29**	76.7%
Precision	96.9%	90.3%	93.3%	80.8%	**95.3%**

(a) Cam0

Predicted True	water	ice	snow	clutter	Recall
water	**211.15**	1.83	2.69	0.00	97.9%
ice	2.10	**27.58**	2.21	0.02	86.4%
snow	1.02	3.60	**94.01**	0.54	94.8%
clutter	0.00	0.02	1.22	**2.82**	69.4%
Precision	98.5%	83.5%	93.9%	83.4%	**95.7%**

(b) Cam1

Table 1. Confusion matrices for the two webcam datasets. Units are millions of pixels, except for precision and recall.

best among the target classes, regarding both recall and precision. For Cam0, recall and precision of all main classes are in the range of 88.3%–98.0%, respectively 90.3%–96.9%. For Cam1, recall and precision of the main classes are 86.4%–97.9%, respectively 83.5%–98.5%. Evidently, the class *ice* is harder to predict than *water* and *snow*, for both data sets. For both Cam1 and Cam2 data sets the recall and precision of the *clutter* class are comparably low. This is mostly due to mistakes on thin structures. We note that the *clutter* class forms only a tiny portion of the pixels, and would be excluded in post-processing (e.g., temporal smoothing) in most practical applications. Somewhat suprisingly, overall accuracy, precision and recall from the low and high resolution streams are comparable. However, for the most challenging period, during freezing, predictions from lower resolution seem to be less stable, see figure 4. Since samples from the freezing period form only a small portion of the data, their higher uncertainty has little impact on the overall numbers. We expect that further reducing resolution, and thus descriptiveness of local texture, will eventually decrease segmentation performance.

4.3 Quantitative Results: Ice On / Ice Off

Freezing and thawing dates are of particular interest for climate monitoring. In this section we seek to exploit temporal redundancy and estimate the daily percentage of ice and snow coverage for the observed water body. Per image, we sum the pixels of each class to obtain the covered area. We then compute the median coverage per class for each day. Finally, the coverage of the water body by *ice*, *snow* and *clutter* (mainly representing man-made structures erected on the ice) are summed. Predictions and ground truth coverage derived from manually labeled segmentations are displayed in figures 4 and 5 for the two cameras. Gaps (marked by red sections) are caused by missing data due to technical problems. For areas where data is available, ground truth is reproduced rather well. For Cam0, an image-wise ice on/ice off classification by thresholding at 50% water yields more than 98% correct predictions, (2 misclassified days with ice coverage near 50% coverage, where minimal differences lead to a flip of the binary prediction). For Cam1, the same threshold classifies all days correctly. We note that true ice-on/ice-off prediction should of course cover the entire lake and account for projective distortion of the lake surface, still the results indicate that an aggregated per-lake analysis will be accurate enough for most applications of interest. Note also, for this evaluation only the test set (25% of all images) was used. Once an operational system is in place, the temporal density will be 4× higher, further increasing robustness.

4.4 Qualitive Results

Figures 6 and 7 show example segmentation results for Cam0 and Cam1, respectively. Column (a) shows the original images, column (b) the corresponding ground truth segmentation. Column (c) shows the automatically generated semantic segmentations, and reliability maps are displayed in column (d). The reliability at a pixel $\mathbf{x}$ is defined as the maximum probability over all classes $r(\mathbf{x}) = max_c(p^c(\mathbf{x}))$.

The first three rows in each of the figures are examples of good segmentation results for interesting, non-trivial input images. As expected, reduced reliability is generally observed near class transitions. There is a tendency for misclassifications to occur in the upper part of the images, presumably due to the loss of high-frequency texture. Even thin structures of the *clutter* class are segmented fairly well. The last row in each figure shows an example where segmentation fails. For failure cases, blocky artifacts appear in the reliability maps, as a result of the tiled processing.

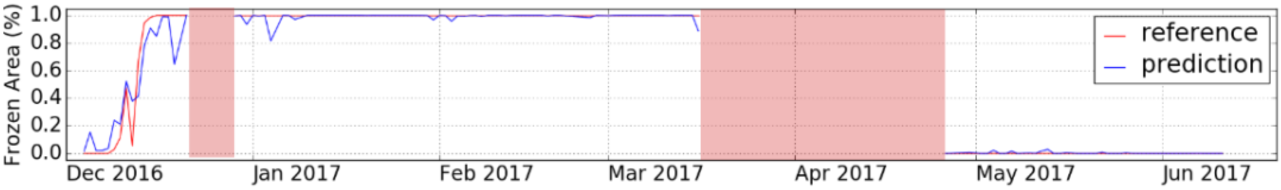


Figure 4. Predicted vs. groundtruth frozen area for Cam0. Red bars indicate periods of data gaps, where no images were stored.

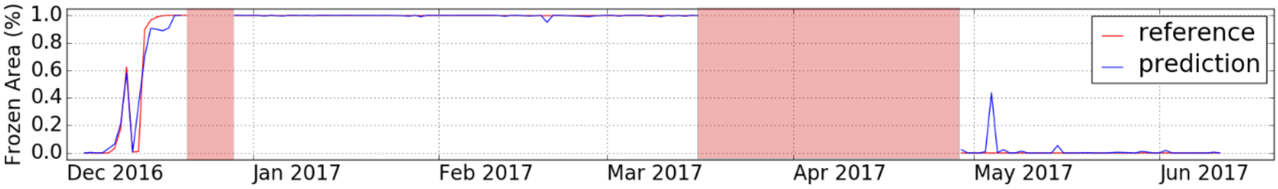


Figure 5. Predicted vs. groundtruth frozen area for Cam1. Red bars indicate periods of data gaps, where no images were stored.

Note that in a number of failure cases even human operators have difficulties, unless they use the temporal context.

5. CONCLUSIONS AND OUTLOOK

In this work, we have investigated the monitoring of lake ice, using webcams instead of traditional remote sensing images as a data source. We have employed a neural network to conventional RGB webcam images to obtain semantic segmentation maps for the lake of St. Moritz. With a class nomenclature of *water*, *ice*, *snow*, *clutter*, we have achieved segmentation accuracies larger than 95% on two different test sequences. We found that among the main target classes, *ice* was the most difficult to predict, but still reached more than 85% recall at more than 80% precision. At the image level, aggregated daily ice-on/ice-off classification by simple thresholding resulted in only two misclassified days over hundreds of images from the winter 2016/2017, both during partial ice coverage near 50%. Overall, we believe that there is large potential to operationally use conventional webcams for lake ice monitoring.

Since images overlap and are captured in rather dense temporal sequences, a future direction of work is to exploit spatial and temporal redundancy to remedy the remaining classification errors. Of particular interest is a more accurate segmentation during the transition periods with partial ice coverage, while stable lake states (water only, full snow or ice coverage) are already classified with very high accuracy. While temporal smoothing appears straight-forward, fusing observations from different cameras requires knowledge of their relative orientation. While stable tie points are hard to find, e.g., after snowfall, one could possibly match silhouettes in mountain areas between images and also to digital elevation models, or match lake borders across cameras. We also plan to carry out experiments to assess the generalization capabilities of already trained networks to new lakes or cameras. Of special interest is the generalization across winters, to simplify long-term observations. To that end we have started to record imagery for the winter 2017/2018.

REFERENCES

Arslan, A. N., Tanis, C. M., Metsmki, S., Aurela, M., Bttcher, K., Linkosalmi, M. and Peltoniemi, M., 2017. Automated webcam monitoring of fractional snow cover in northern boreal conditions. *Geosciences* 7(3), pp. 55.

Badrinarayanan, V., Kendall, A. and Cipolla, R., 2015. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*.

Bogdanov, A. V., Sandven, S., Johannessen, O. M., Alexandrov, V. Y. and Bobylev, L. P., 2005. Multisensor approach to automated classification of sea ice image data. *IEEE Transactions on Geoscience and Remote Sensing* 43(7), pp. 1648–1664.

Bothmann, L., Menzel, A., Menze, B. H., Schunk, C. and Kauermann, G., 2017. Automated processing of webcam images for phenological classification. *PloS one* pp. 1–23.

Bradley, E. S. and Clarke, K. C., 2011. Outdoor webcams as geospatial sensor networks: Challenges, issues and opportunities. *Cartography and Geographic Information Science* 38(1), pp. 3–19.

Brown, L. C. and Duguay, C. R., 2010. The response and role of ice cover in lake-climate interactions. *Progress in Physical Geography* 34(5), pp. 671–704.

Duguay, C., Bernier, M., Gauthier, Y. and Kouraev, A., 2015. *Remote sensing of lake and river ice*. Wiley-Blackwell, Oxford, pp. 273–306.

Dutta, A., Gupta, A. and Zissermann, A., 2016. VGG image annotator (VIA). http://www.robots.ox.ac.uk/ vgg/software/via/. Accessed: 11.01.2018.

Fedorov, R., Camerada, A., Fraternali, P. and Tagliasacchi, M., 2016. Estimating snow cover from publicly available images. *IEEE Transactions on Multimedia* pp. 1187–1200.

Glorot, X., Bordes, A. and Bengio, Y., 2011. Deep sparse rectifier neural networks. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323.

He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep residual learning for image recognition. *CoRR*.

Huang, G., Liu, Z. and Weinberger, K. Q., 2016. Densely connected convolutional networks. *CoRR*.

Ioffe, S. and Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the International Conference on Machine Learning*, pp. 448–456.

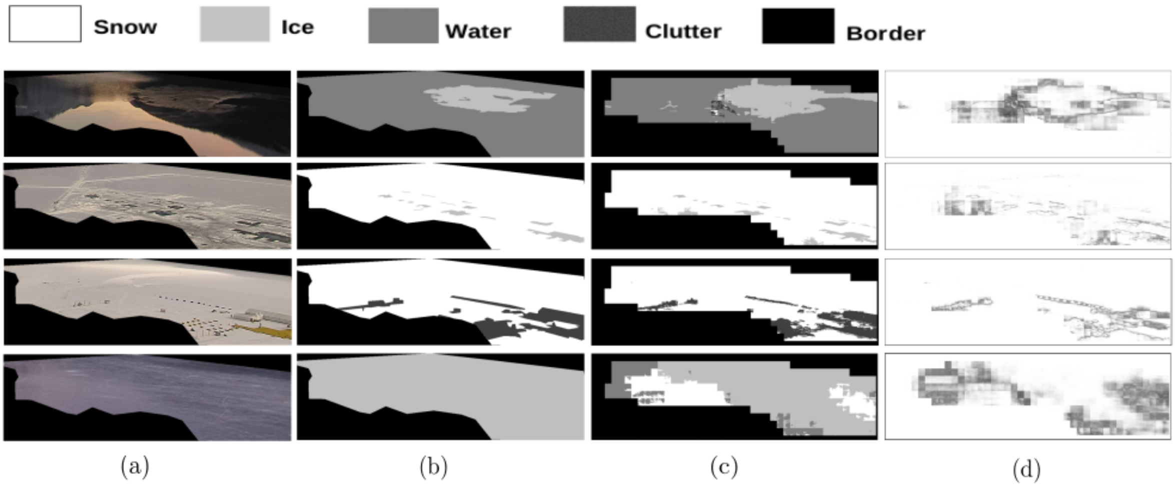


Figure 6. Semantic Segmentations for Cam0. (a): RGB input; (b): manual segmentations; (c): predicted segmentations; (d): reliability maps.

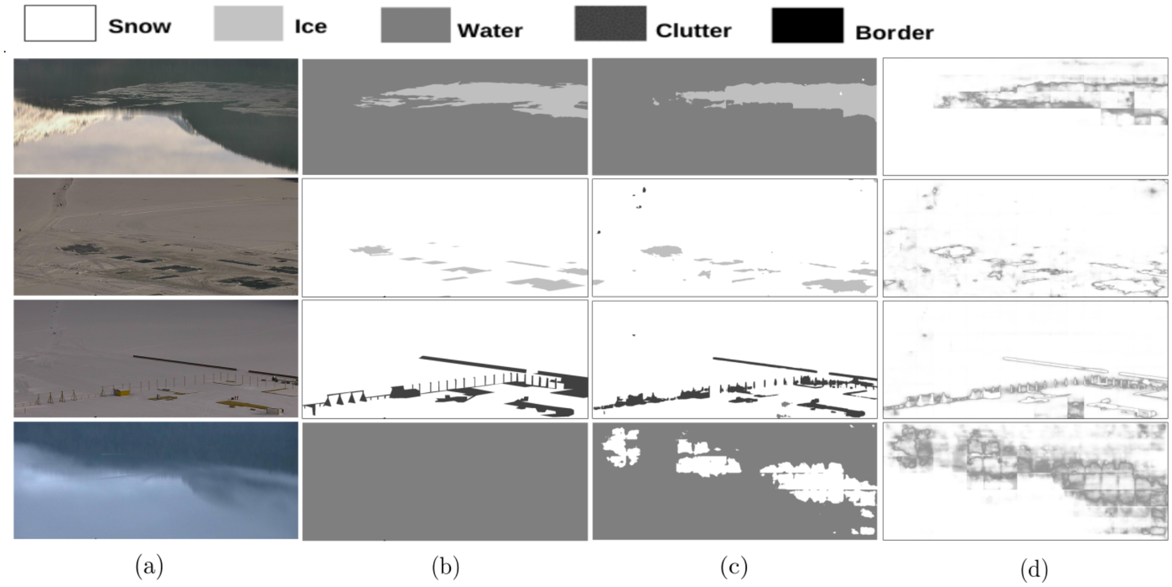


Figure 7. Semantic Segmentations for Cam1. (a): RGB input; (b): manual segmentations; (c): predicted segmentations; (d): reliability maps.

Jacobs, N., Burgin, W., Fridrich, N., Abrams, A., Miskell, K., Braswell, B. H., Richardson, A. D. and Pless, R., 2009. The global network of outdoor webcams: properties and applications. In: *Proceedings of ACM SIGSPATIAL*, pp. 111–120.

Jacobs, N., Roman, N. and Pless, R., 2007. Consistent temporal variations in many outdoor scenes. In: *CVPR'07*, pp. 1–6.

Jégou, S., Drozdzal, M., Vázquez, D., Romero, A. and Bengio, Y., 2016. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *CoRR*.

Latifovic, R. and Pouliot, D., 2007. Analysis of climate change impacts on lake ice phenology in canada using the historical satellite data record. *Remote Sensing of Environment* 106(4), pp. 492 – 507.

Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *CVPR 2015*, pp. 3431–3440.

Richardson, A. D., 2015. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery: A new, publicly-available dataset. *AGU Fall Meeting Abstracts*.

Richardson, A. D., Jenkins, J. P., Braswell, B. H., Hollinger, D. Y., Ollinger, S. V. and Smith, M.-L., 2007. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia* 152(2), pp. 323–334.

Robertson, D. M., Ragotzkie, R. A. and Magnuson, J. J., 1992. Lake ice records used to detect historical and future climatic changes. *Climatic Change* 21(4), pp. 407–427.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *CoRR*.

Rüfenacht, D., Brown, M., Beutel, J. and Süsstrunk, S., 2014. Temporally consistent snow cover estimation from noisy, irregularly sampled measurements. In: *Computer Vision Theory and Applications (VISAPP)*, pp. 275–283.

Salvatori, R., Plini, P., Giusto, M., Valt, M., Salzano, R., Montagnoli, M., Cagnati, A., Crepaz, G. and Sigismondi, D., 2011. Snow cover monitoring with images from digital camera systems. *Ital. J. Remote Sens.* 43(2), pp. 137–145.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research* pp. 1929–1958.

Sutskever, I., Martens, J., Dahl, G. and Hinton, G., 2013. On the importance of initialization and momentum in deep learning. In: *Proceedings of the International conference on machine learning*, pp. 1139–1147.

Sütterlin, M., Duguay-Tetzlaff, A. and Wunderle, S., 2017. Toward a Lake Ice Phenology Derived from VIIRS Data. In: *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, p. 8907.

Tom, M., Lanaras, C., Baltsavias, E. and Schindler, K., 2017. Ice detection in swiss lakes using modis data. https://www.ethz.ch/content/dam/ethz/special-interest/baug/igp/photogrammetry-remote-sensing-dam/documents/pdf/Papers/.

Xie, S., Girshick, R. B., Dollár, P., Tu, Z. and He, K., 2016. Aggregated residual transformations for deep neural networks. *CoRR*.