CNN Classification of Wet Snow By Physical Snowpack Model Labeling

Matthieu Gallet, Abdourrahmane Atto, Emmanuel Trouvé, Fatima Karbou

CNN CLASSIFICATION OF WET SNOW BY PHYSICAL SNOWPACK MODEL LABELING

Matthieu Gallet[†], *Abdourrahmane Atto*[†], *Emmanuel Trouvé*[†], *Fatima Karbou*[‡]

[†] : Université Savoie Mont Blanc, LISTIC, Annecy, France
Contact : {firstname.lastname}@univ-smb.fr
[‡] : Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS
CNRM, Centre d'Études de la Neige, 38000 Grenoble, France, Contact : fatima.karbou@meteo.fr

ABSTRACT

We propose a new approach for wet snow extent mapping in Synthetic Aperture Radar (SAR) images by using a convolutional neural network (CNN) designed to learn with respect to snowpack outputs from the state-of-the-art snow model Crocus. The CNN was trained to classify the wet snow conditions based on features extracted from the SAR images, using both the VV,VH channel and the ratio between these channels and those of a reference image in summer. One of the key points of this work is the comprehensive comparison we have made between the performance of the CNN method and other advanced statistical methods. We found that the CNN was able to achieve good accuracy in wet snow classification, and giving a complementary vision of the solutions obtained by other machine learning algorithms such as the Random Forest classifier. The results of this study demonstrate the potential of using CNNs and SAR images for wet snow classification and highlight the importance of using physical information model for training machine learning models in snow state identification, a domain where collecting ground truth is intricate due to the complexity of the snowpack moisture measurement systems.

Index Terms— SAR, Wet Snow, Classification, Convolutional Neural Networks

1. INTRODUCTION

The identification of snow states is essential for many applications including understanding and management of water resources on our planet. With the use of multi-spectral or SAR data, it has become feasible to observe snow cover on a global scale. In this context, SAR images are particularly interesting because of their ability to observe day and night whatever the meteorological conditions. Many studies have focused on the use of backscattering signal to detect, map or quantify snow [1]. More recently some works have integrated computer vision methods to perform these classification or segmentation tasks [2]. The Sentinel-1 mission open data has made it possible to monitor snow conditions over large areas with a revisit time of a few days. The characterization of wet snow remains complex [3] and one of the first studies [4] proposed the mapping of wet snow by a thresholding on the decibel data using one or two polarimetric channels [5]. Other studies have used machine-learning methods using a large number of auxiliary data [6].

In this work, we propose first, a procedure for labeling SAR data with respect to simulation outputs from the snow evolution model CROCUS [7],[8]. The labeling procedure is given in Section 2. In a second step, we propose in Section 3, to exploit the recent advances in computer vision and deep-learning networks to realize the binary classification task of wet snow. We propose to compare a CNN architecture with a low number of layers in order to keep the execution speed and computational frugality, to a more traditional machine-learning algorithm. The results are evaluated at the scale of a complete massif with standard machine-learning metrics. We focus on a portion of the test area on one date to study the wet snow extension maps. These results are compared to the optical and radar products provided by Copernicus, respectively the Fractional Snow Cover (FSC) obtained by Sentinel 2 with a resolution of 20m and the SAR Wet Snow (SWS) obtained using Sentinel 1 with a resolution of 60m.

2. PHYSICAL MODEL BASED SAR LABELING

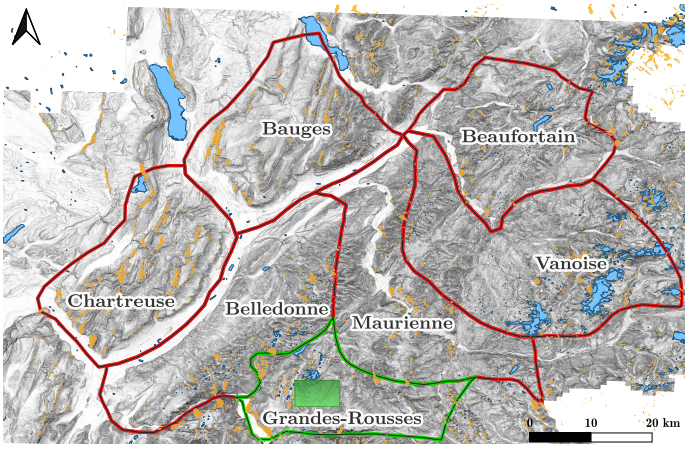

Fig. 1: Training areas in red, test areas in green and validation areas (green rectangle) located in the north of the French Alps

Labels	X_{train}	X_{test}
Wet	16560	2099
Not Wet	470597	2099

Table 1: Number of samples distributed in the classes for the training (unbalanced) and test (balanced) datasets

We consider in this work the Sentinel 1 GRD data between July 2020 and August 2021, that is 69 images at 20m resolution. These data are composed by two polarimetric channels VV and VH. We rely on the CNES computing facilities to pre-process Sentinel-1 data; the pre-processing includes thermal noise removal, speckle filtering, radiometric calibration and terrain correction using the SRTM DEM (30 m). In this study, we use only the ascending data. Indeed, the acquisition time of the ascending data is 17:30 UTC. At this time of the day, the snowpack has warmed up all day, allowing the detection of larger areas of wet snow. We also use the ratio between the current date and a reference date that we select in summer when the snow extent is minimal. The reference date is August 09, 2020. It is a generalization of Nagler's method [4], where a static thresholding is applied, in contrast to our approach where thresholding can be adjusted dynamically.

The study is carried out on a portion of the 31TGL tile, which is part of the Sentinel-2 tiling system and the Military Grid Reference System, located in the Northern French Alps. As illustrated in Figure 1, this area includes 6 complete massifs to realize the training dataset (in red) and 1 massif for the test dataset (in green). We perform a paving of each massif and select windows of size 16 by 16 pixels distant from 80 pixels for each date, in order to avoid a too important correlation between the samples. The size of this analyzing window allows a sufficiently deep decomposition by the joint use of convolutions and max-pooling operations. We then perform the normalization between 0 and 1 of all samples, using the dynamics of the entire time series. Finally the dataset has the following dimension : $N \times 16 \times 16 \times 4$ with N the number of samples in training or in test.

For labeling purposes, [9] provides a definition of wet snow based on the temperature component of the snowpack, and considers that if it is higher than zero degrees Celsius we are in presence of wet snow. In our study we rely on snowpack reanalyses from the snow model Crocus [8]. Snowpack simulations were carried out using the detailed snow cover model Crocus coupled with the ISBA land surface model within the SURFEX (EXternalized SURFace) simulation platform.The model chains (snow/soil) describe the evolution of the physical properties of the snowpack, its stratigraphy (up to 50 layers) and the underlying ground, under meteorological forcing data. We consider two CROCUS variables of the snowpack simulated à 6 PM: the minimum temperature T_n (in degrees Celsius) and the snow height H_s (in meter). We use the following labeling rule:

$$\text{Wet Snow} = T_n > 0 \text{ and } H_s > 0.4, \quad (1)$$

In addition of the rule on the temperature of the snowpack given by [9], we use a constraint on the snow depth H_s. This requirement ensures that the radar signal comes primarily from the presence of wet snow, rather than from the ground response under a thin layer of snow or from rocky outcrops with minimal snow cover.

The wet snow identification criteria given by Eq.1 makes labeling of large areas seen by SAR straightforward, whatever the acquisition time. However, the spatial resolution of CROCUS is low, with a discretization of approximately 20 degrees in slope, 300 meters in altitude, and 22.5 degrees in orientation, representing sometimes important zones within a massif, which are locally heterogeneous. Moreover this truth can be subject to model errors, not present when using specific measurement stations. Nevertheless, the use of a physical model for supervised learning allows us to obtain a large and a varied training dataset, representative of the inter and intra-massif topological variabilities, allowing a sufficiently general representation for its application to other geographical areas. The created labeled dataset is divided into two classes: *wet* and *not wet*, described in Table 1. The *not wet* class contains the set of snow-free or dry-snow samples. The class *wet* contains the samples likely to be wet snow under the constraints of the CROCUS model estimation.

3. METHODOLOGY

Based on the study of CNN architectures for SAR proposed by [10], we have developed a shallow CNN network as illustrated in Figure 2 with 5497 parameters to learn. The idea is to exploit the specific characteristics of CNNs, namely the learning of convolutional layers used for feature extraction. This is achieved in our case by keeping only 3 blocks (in blue and green in the figure) consisting in cascade of a 2D convolution layer, a max-pooling (allowing to reduce the number of parameters) and a ReLu activation. The second part of the network (in orange in the figure) focuses on learning the combinations of these decompositions by fully connected layers. Two dense layers make up this part in order to progressively reduce the number of parameters to the number of classes. To avoid overfitting we added a drop out layer which leaves only 25% of the values of the previous layer randomly during learning. The representation of the dataset is important for this type of approach and in order to avoid a learning bias, we perform a balancing of the training dataset between the classes. This can be done by random sub-sampling to balance the majority *not wet* class. This technique allows us to have an equal number of samples in both classes, but if there is a significant imbalance, it significantly reduces the majority class, going in our case from 474K samples to 33K.

To study the whole dataset, we perform a K-fold study for the training dataset for the majority class case: *not wet*. This last one is randomly divided in K sub-parts of size equal to the size of minority class *wet*, in this way we can study the

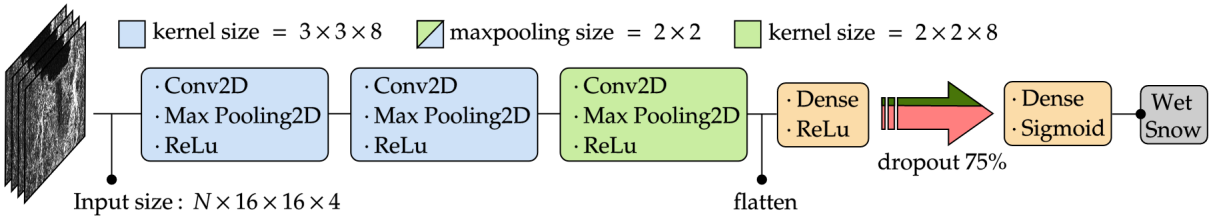

Fig. 2: CNN architecture

totality of the created dataset and the variability of the modality according to the full dataset. In our case, $K = 29$. We give the performance of the classifier on the test dataset which will be balanced and this for each training fold, with the following metrics: F1 score, Cohen's kappa (κ), AUC. We also give a mean confusion matrix for the training set calculated on the balanced test dataset. We introduce two additional indicators: the Best Accuracy-ROC (BAROC) and the Constant False Positive Rate-ROC (FCROC). By looking at the output of the unthresholded model, i.e. directly the probability of belonging to the *wet* class, we study which threshold value on these probabilities yields respectively the best accuracy and a constant false positive rate. We set the false positive rate at 5%. Thus we apply this new thresholding to refine the results provided by the CNN according to our use. Keeping the false positive rate low and constant ensures that the amount of *wet* or *not wet* snow is not overestimated. We compare the results obtained by the CNN network with the use of a Random Forest classifier on vectorized data used in a more complex way in [6]. Finally, we present a map obtained with the two thresholds (BAROC and FCROC at 5%). The code is available online: `https://github.com/Matthieu-Gallet/CNN-WetSnow-CROCUS`.

4. RESULTS

Quantitative evaluation: The comparison between the proposed network and the random forest approach is given in Table 2. We observe that whatever the metric, the CNN network outperforms the Random Forest. The most important difference is the kappa coefficient. We can note a more important variability on the side of the CNN, which can be explained by the facility of the latter to overfit in certain folds, demonstrating the need to consider the data in their entirety and thus to take into account the balancing of the dataset directly in the cost function. The confusion matrices for the BAROC and FCROC thresholds given in Table 3 show, in addition to the superior performances on the positive and negative true number obtained with the CNN, a better balance between the false detections related to the *wet* and *non-wet* labels than the false detections obtained by the random forest. It is interesting to

	F1	AUC	κ	BAROC	FCROC
CNN	76.3 +/- 1.8	83.9 +/- 0.6	52.8 +/- 3.2	78.1 +/- 0.6	69.3 +/- 0.9
RF	75.2 +/- 0.2	83.6 +/- 0.1	50.6 +/- 0.4	77.0 +/- 0.2	68.8 +/- 0.4

Table 2: Comparison of results between a CNN and a Random Forest classifier (x100)

True		Predict nw	Predict nw	Predict w	Predict w
nw		40.7	**41.4**	9.3	**8.6**
w		**12.6**	14.5	**37.4**	35.5

(a) BAROC

True		Predict nw	Predict nw	Predict w	Predict w
nw		47.5	**47.6**	2.5	**2.4**
w		**28.2**	29.1	**21.8**	20.9

(b) FCROC

Table 3: Confusion matrices for BAROC and FCROC thresholds for the CNN (black) and the Random Forest (red) in % for the *not wet* (nw) and wet (w) classes.

note that the thresholds give a specific information: while the BAROC thresholding favors a high global accuracy and a balanced false detection rate between the 2 classes, the FRCOC thresholding allows to have a low false detection rate for the *not wet* class, at the expense of a higher false detection rate for the *wet* class. Two things can be concluded from this, the first is that the two pieces of information can be complementary to each other in order to estimate the amount of wet snow. Second interesting aspect is that the accuracy of convolutional networks is limited to some extent, by the complexity of differentiating the two classes. Note that a physical model error or bias can generate a systematic misclassification, and therefore there is a need for some measure of imperfect labelling.

Visual comparison: Figure 3 presents the study area specified in green in Figure 1 in the melt period on March 31st, 2021. Figure 3.**B** corresponds to the logical combination of the two classifiers (RF and CNN) using BAROC thresholding, on which the FSC product was added in Figure 3.**C**, the SWS product in Figure 3.**D** and the CROCUS map with the condition in (1) in Figure 3.**E**. We notice that the four maps are similar from a global point of view. More in detail, in figure 3.**B** we notice that the 2 classifiers (RF and CNN) share large areas. The main distinction between the two is the presence of wider edges on the areas detected with the random forest. Those only detected by the CNN are few and very localized. On March 31, 2021, the snowpack has already begun to moisten, allowing us to compare the results obtained with the optical snow cover product (3.**C**). The two maps are very close visually. One can see a large area at the top of the image, which is seen as snow but is not considered as wet in the results. Otherwise the boundaries between wet snow and the rest, especially in the lower part of the image, are close to those detected by FSC. The major difference is the diagonal in the right part of the image, which does not show snow in the FSC but is detected by the RF as wet snow. The RF seems to overestimate the areas. We can only speculate that the response of this area may be due to a larger roughness due to the presence of rock and the outcrop of rock or an area that

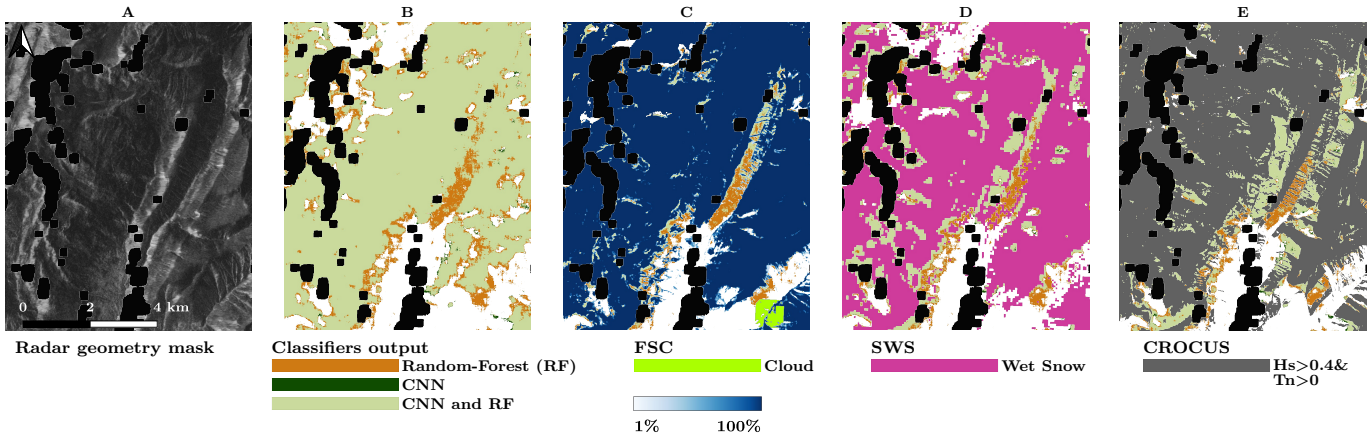

Fig. 3: Results for the validation area in 1 on March 31st, 2022. Layover and shadow are masked in black. **A**. Sentinel 1 image. **B**. Overlaid results of the BAROC (orange) and FCROC (red) methods. **C**. Overlay on the results of the FSC product. **D**. Overlay on the results of the SWS product. **E**. Overlay on the results of the CROCUS model under the rule given in (1).

can be almost considered as a shadow-type distortion. This diagonal is partly found in the SWS wet snow product (3.**D**). The higher resolution at which the network works makes it possible to obtain on other areas (bottom left) a fairly accurate cutout of the mixture between the presence of *wet* snow and *not wet* snow compared to the FSC and SWS product. Finally the last map shows the overlay of our results with the the labeling rule obtained by the CROCUS model. As pointed out, the model is subject to imperfections. Some parts of the image are given as wet snow by the CROCUS model but are not seen as having snow by the FSC. However it is interesting to note that the CROCUS model has a lower spatial extension than our results and the SWS product (more precisely on the right side of the image). One can assume that this discrepancy comes from the sensitivity of the methods to wet snow for all considered snow heights and not only above 40 cm.

5. CONCLUSIONS

In this paper we have illustrated the ability of CNN networks to detect wet snow from a SAR image dataset with a resolution of 20m per pixel and labeled using simulations from the physical model of the snowpack: CROCUS. This development has been compared from a machine learning point of view to a more conventional classifier: the random forest, against which it has demonstrated a slight superiority on the studied metrics. Moreover, we were able to compare directly on a study area at a given date the wet snow mapping obtained by this approach with the existing Copernicus products (FSC and SWS). This evaluation allowed us to assess the relative good correspondence between the 3 estimates.The limitations of this method are related to 2 major points: the reliability of the physical model used to determine the labels of the samples and the balance of the dataset that could be directly taken into account in the cost function.

6. REFERENCES

[1] Fatima Karbou, Isabelle Gouttevin, and Philippe Durand, "Spatial and temporal variability of wet snow in the french mountains using a color-space based segmentation technique on Sentinel-1 SAR images," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 5586–5588.

[2] Rahul Nijhawan, Josodhir Das, and Balasubramanian Raman, "A hybrid of deep learning and hand-crafted features based approach for snow cover mapping," *International Journal of Remote Sensing*, vol. 40, no. 2, pp. 759–773, Number: 2.

[3] Carlo Marin, Giacomo Bertoldi, Valentina Premier, Mattia Callegari, Christian Brida, Kerstin Hürkamp, Jochen Tschiersch, Marc Zebisch, and Claudia Notarnicola, "Use of Sentinel-1 radar observations to evaluate snowmelt dynamics in alpine regions," vol. 14, no. 3, pp. 935–956, Number: 3.

[4] T. Nagler and H. Rott, "Retrieval of wet snow by means of multitemporal SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 754–765.

[5] Thomas Nagler, Helmut Rott, Elisabeth Ripper, Gabriele Bippus, and Markus Hetzenecker, "Advancements for snowmelt monitoring by means of Sentinel-1 SAR," vol. 8, no. 4, pp. 348, Number: 4.

[6] Tsai, Dietz, Oppelt, and Kuenzer, "Wet and dry snow detection using Sentinel-1 SAR data for mountainous areas with a machine learning technique," *Remote Sensing*, vol. 11, no. 8, pp. 895, Number: 8.

[7] Eric Brun, Paul David, Marcel Sudul, and Gilles Brunot, "A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting," *Journal of Glaciology*, vol. 38, pp. 13 – 22, 1992.

[8] V. Vionnet, E. Brun, S. Morin, A. Boone, S. Faroux, P. Le Moigne, E. Martin, and J.-M. Willemet, "The detailed snowpack scheme Crocus and its implementation in surfex v7.2," *Geoscientific Model Development*, vol. 5, no. 3, pp. 773–791, 2012.

[9] Ya-Lun S. Tsai, Andreas Dietz, Natascha Oppelt, and Claudia Kuenzer, "Remote sensing of snow cover using spaceborne SAR: A review," *Remote Sensing*, vol. 11, no. 12, pp. 1456, Number: 12.

[10] Hemani Parikh, Samir Patel, and Vibha Patel, "Classification of SAR and PolSAR images using deep learning: a review," *International Journal of Image and Data Fusion*, vol. 11, no. 1, pp. 1–32, Number: 1.