Image-Based Ordinal Regression of Snow Depth: A deep learning approach.

EGU
ITS4.6/NH6.7
Data Science and Machine Learning for Natural Hazards and Seismology

Pierre Lepetit (Météo-France, LATMOS)

Cécile Mallet (LATMOS), Laurent Barthès (LATMOS)

Observation of the snow cover : why and how.

In french plains, snow accumulation is a rare, local and hardly predictible phenomenon which could have heavy consequences on the road transports. Nowdays, monitoring of snow depth mainly relies on human observation, satellite teledetection and local specific sensors.

fig.1. Traffic congestion due to snow accumulation on A10 highway in Ile-de-France. L'EXPRESS, 8/12/2010.

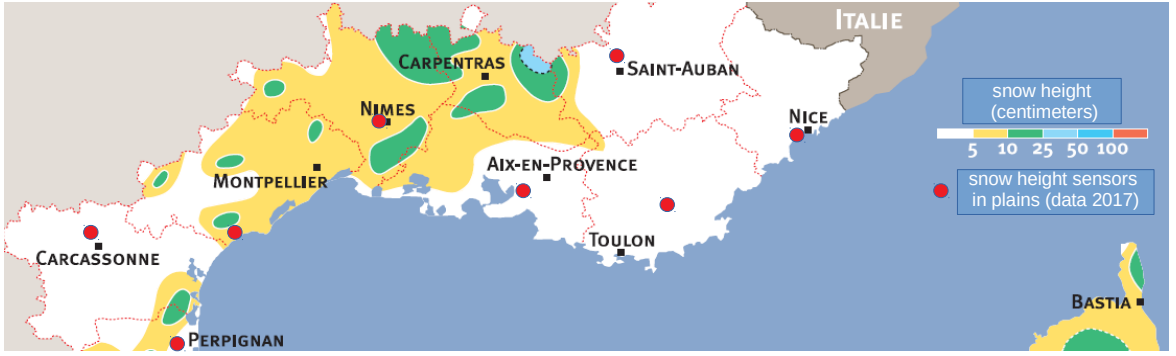


fig.2, modified from [1] : example of snow accumulation on mediterranean plains after the snowfall event (28-31/12/1996). Meteo-France automatic sensors don't allow an accurate monitoring of the snow height evolution.

On the one hand, human observation and local sensors are expensive and teledection through snowfall is challenging. On the other hand, webcam data became ubiquitous and progress has been made in image-based estimation of meteorological parameters [8, 11, 6, 12, 17, 9, 5].

Here, we present an approach of the **quantitative characterization** of the snow cover based on webcam images. We train a Machine Learning model to yield a quantitive index correlated with the snow depth measured by specific sensors.

Scope of the study

To train and test our model on relevant images, we built our datasets from daytime and nighttime images taken around starts of snow events. We limited the corpus to RGB images taken by road webcams, because they form the main part of the available webcam data, and because they contain critical informations for operational meteorology.

fig.3. Example of webcam images. **Scenes** are **urban**, **rura**l. They contain avenues, highways, field roads, bridges, parkings, etc, under natural or artificial lightings.

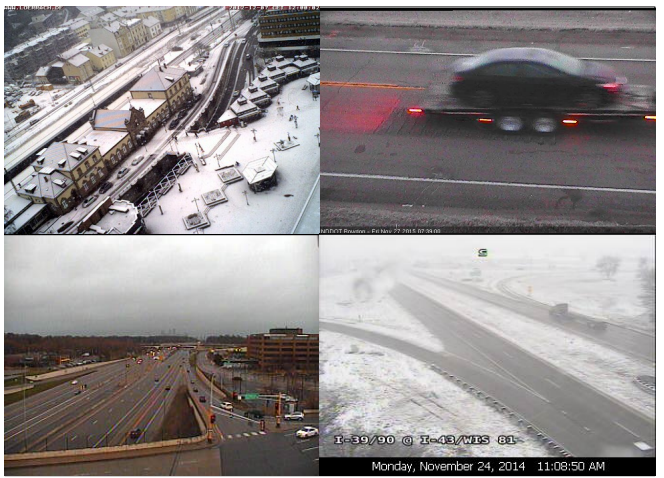


fig.4. Road webcam **orientations** are heterogene. Direction varies from the horizontal to bird's eye views. Optics and numerical post-processing also strongly vary.

fig.5. Webcam images are often corrupted by **natural masks** (snowflakes, droplets, filth) and numerical overprints (text, logos, marks, etc).

On these images, the difficulties lie in the diversity of the road scenes, the heterogeneity of the devices and in the noises due to adverse weather conditions (see *fig.5* - 15 % of our images are corrupted).
Machine Learning approaches based on wide datasets may yield interesting results while facing these difficulties. But in our case, there is an obstacle : there's a lack of accurate quantitative labels.

Previous work

On the detection and the quantification of the snow:

Previous studies on image-based snow charecterization deal with classification tasks (e.g. snow vs. no snow) [3, 4, 5]. In these studies, the datasets are not conceived to analyse the snow growth during snowfalls. Images are generally taken around midday [3] and by good weather [3,4,5].

Moreover, to our knowledge, there has **been no attempt to rank snow depth** by machine learning.

Fig 6. images taken from [3] (first raw, phenocam network) and [5] (second raw,, EC1M dataset).

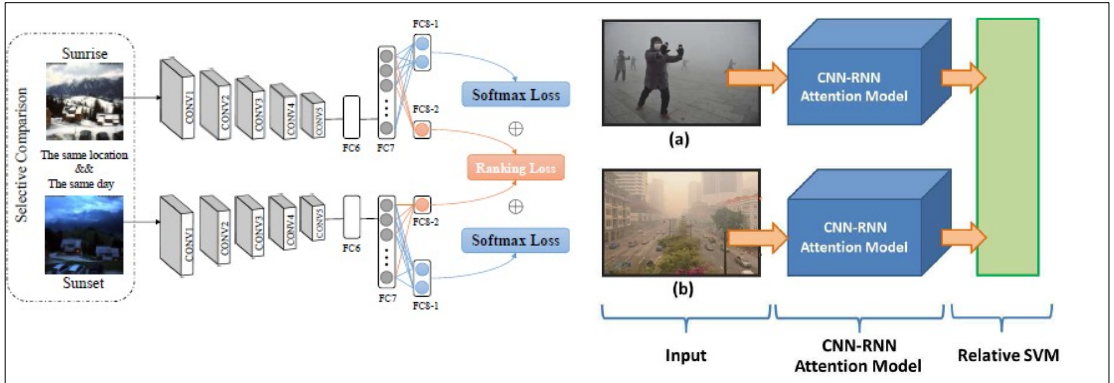


Fig 7. taken from [9] (rights) and [10] (lefts). [9] estimates temperatures and [10], optical range. They both use pairwise learning. [9] trains CNN in an end-to-end siamese mode.

On image-based weather estimation :

- End-to-end deep learning approaches perform better, as in other domains of SUN [6,7,9].
- Pairwise learning with siamese networks is efficient for both classification and regression tasks [9,10,13].
- Weak supervision with handcrafted binary comparisons (ordinal labels) mitigates the lack of quantitative labels [10].

We also use ordinal regression to yield a quantitative index. Our approach is based on an end-to-end ranking model. The training phase relies on a directed graph built from handcrafted ordinal labels.

Building the datasets : **three sources** of webcam images

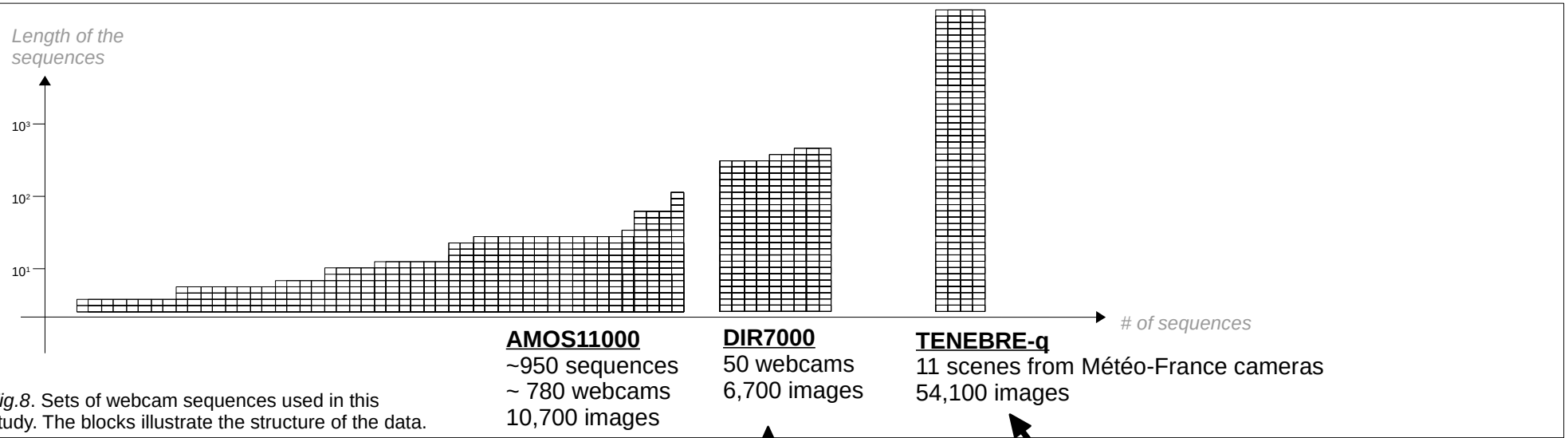


Fig.8. Sets of webcam sequences used in this study. The blocks illustrate the structure of the data.

Collection of AMOS11000 *:*

AMOS [2] is a large database of worldwide webcam images. To sample images of snow events, geo-stamp and Time-stamp of AMOS webcams have been cross-checked with the height of snow from the ERA-5 reanalysis [20].
A manual subsampling has then been made to avoid redundancy, dark nights and strongly corrupted images.

Collection of DIR7000 *:*

The webcam network of the Direction Interdépartementale des Routes was archived during the snow event of the 23-26/01/2019. The time-step is generally finer than for AMOS sequences and the growth of the snow cover is better sampled.

Collection of TENEBRE-q :

The webcams of the TENEBRE network, owned by Météo-France, are hosted in weather stations.
Their archives have been resampled. All the images associated with snow, bad visibility and rain have been kept.
Our TENEBRE-q dataset was then completed by 11000 images taken by good weather (~1000 by scene).

Labelling of **AMOSDIR1800**, **TENEBRE-q** and **TENEBRE-h**

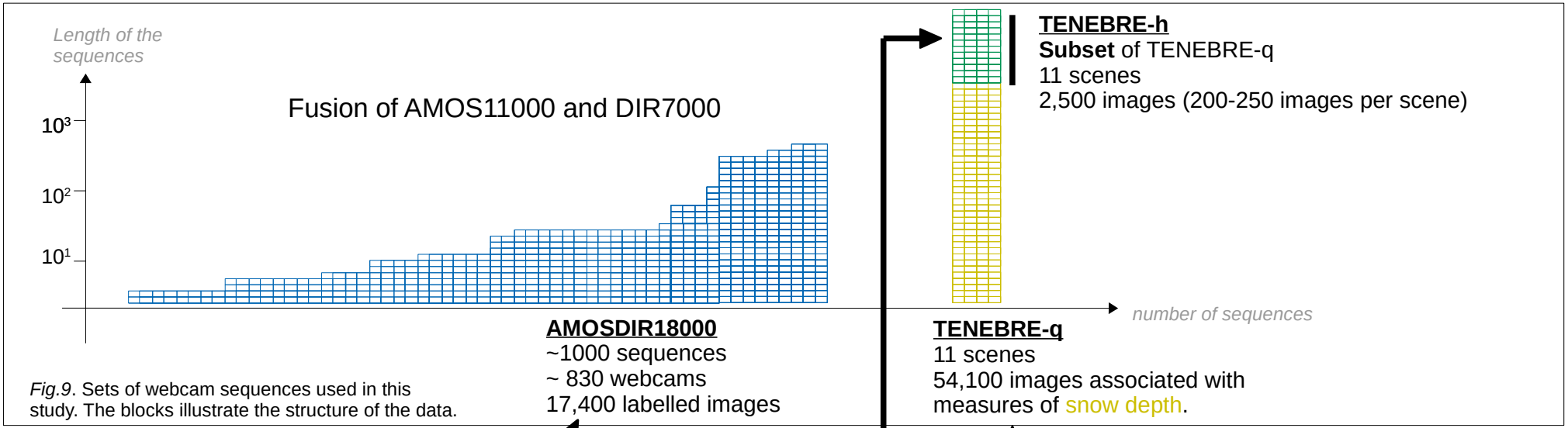


Fig.9. Sets of webcam sequences used in this study. The blocks illustrate the structure of the data.

Handcrafted labelling of **AMOSDIR18000** (and **TENEBRE-h**) **:**

Qualitative labels : 4 levels of snow cover
- 0 : « no snow » (20 % of the dataset)
- 1 : « snow settles on the ground » (40%)
- 2 : « snow settles on the road » (30%)
- 3 : « ground and road are totally covered» (10%)

Ordinal labels : 10,400 pairs of **consecutive** images
- Labels « > » or « < » : 6,500 pairs
- Label « = » : 3,900 pairs

Instrumental labelling of **TENEBRE-q :**

Labels come from snow-depth sensors (JENOPTIK SHM30 and APICAL TLN35R) colocalized with the webcams of TENEBRE. Values are given in centimeters with an incertitude of ± 0.5 cm.

Dual labelling of **TENEBRE-h :**

A subset of TENEBRE-q has been manually labelled. These images have hence both instrumental and handcrafted labels.
It allowed in particular to **check the quality and the coherence** of both modes of labelling.

From hancrafted labels to directed graphs

We apply rules to convert ordinal and qualitative labels into partial orders on image sets.
These partial orders are stocked in four **D**irected **G**raphs : **DG**$_{\text{train-o}}$, **DG**$_{\text{train-q}}$, **DG**$_{\text{val}}$ and **DG**$_{\text{test}}$.

Rules for ordinal labels :

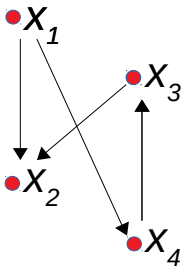
i) Label conversion :

« image x_1 > *image* x_2 » → *new* edge (x_1 , x_2)

« $x_3 < x_4$ » → *new* edge (x_4 , x_3)

« $x_2 = x_4$ » → *new* edges :

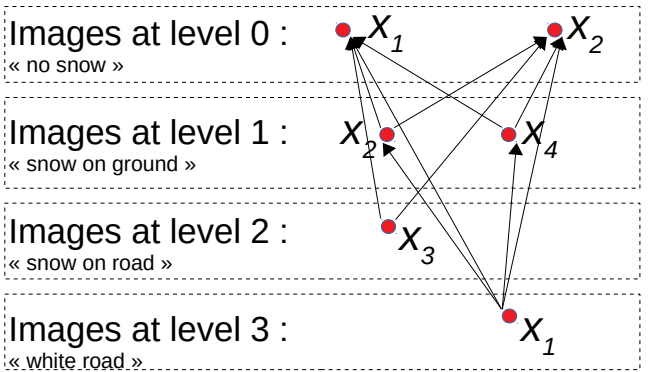
(x_1 , x_4) , (x_3 , x_2)



ii) Take the transitive closure

Rules for qualitative labels :

i) level 0 < levels 1 ; 2 ; and 3. level 1 < level 3.



ii) keep randomly 100 edges max. by webcam sequence

DG$_{\text{train-q}}$	**DG**$_{\text{val}}$	**DG**$_{\text{test}}$	**DG**$_{\text{train-q}}$
14.300 nodes	**1,700** nodes	**2,500** nodes	**14,300 nodes**
28,000 edges	2,500 edges	1,900 edges	61,600 edges

AMOSDIR18000 is split in **training** webcam sequences (80%) and **validation** webcam sequences (20%)

DG$_{\text{train-o}}$, and **Dg**$_{\text{train-q}}$, are made with labelled images of the train sequences

DG$_{\text{val}}$ is made with the validation sequences

DG$_{\text{test}}$ is made with images of **TENEBRE-h**

Training along the graphs

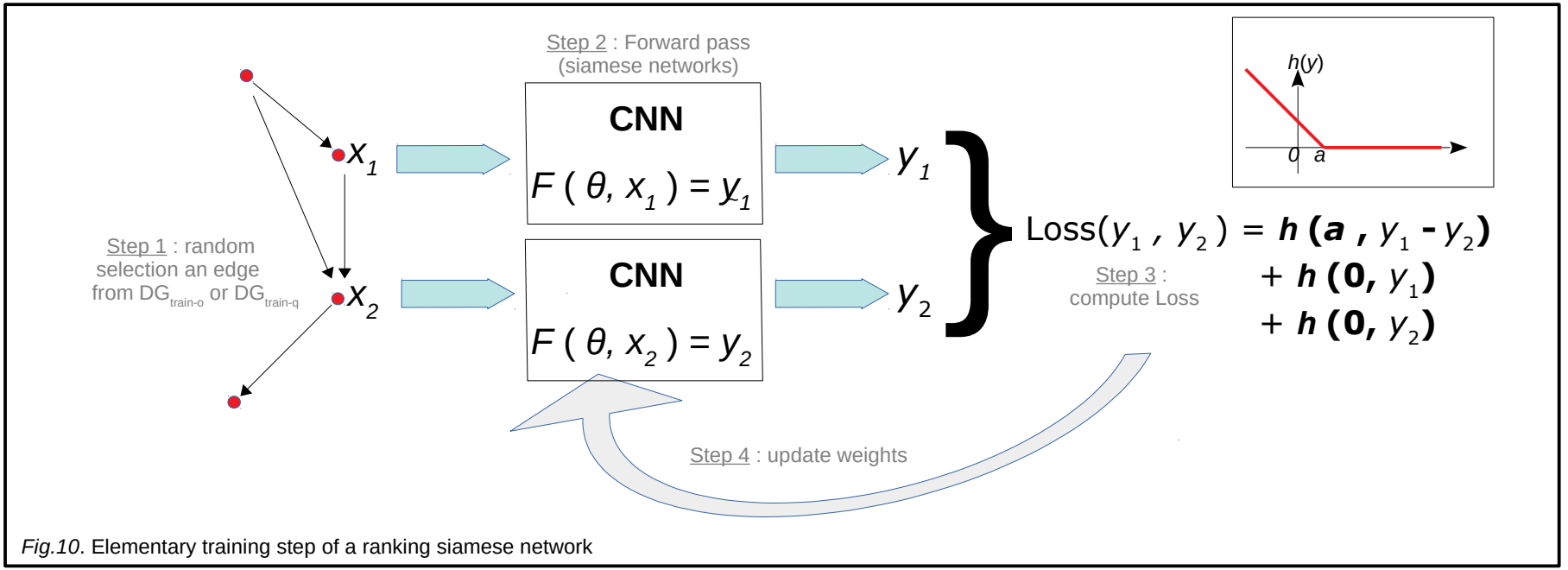


Fig.10. Elementary training step of a ranking siamese network

Edge selection :

- At each step, edges are selected :
  - from $DG_{train\text{-}o}$ with a proba. of p_o
  - from $DG_{train\text{-}q}$ with a proba. of $p_q = 1 - p_o$
- Inside $DG_{train\text{-}o}$, a weighted sampling is used to favour edges between small sequences.

Models :

Two kinds of architectures were tested : Resnet50 [14] and VGG16 [15]. The softmax layer is replaced by a one-dimensional output layer.

Loss Function:

- Built on Ranking Hinge Losses
- The first member penalizes miss-ordered outputs.
- The other terms force the output positivity.

Other Training parameters: Classical data-augmentation tools of were applied. M*ini-batch count 64 cropped images. The gradient descent is optimized through the ADAM algorithm [23] with a starting learning rate of 0.001.*

Results on ranking tasks and model selection

Pretrained models :	VGG16 (places365)	VGG16 (imagenet)	Resnet50 (places365)	Resnet50 (imagenet)
validation (best accuracy)	83.6 %	82.1 %	87.3 %	88,3 %
test (accuracy)	69.9 %	73.3 %	74,2 %	74,8 %

Tab.1. Performances of pretrained CNN

During the validation and the test phases, all the edges of DG_{val} and DG_{test} are browsed. The accuracy is the fraction of the output pairs that are correctly ordered.
We stress the fact that **the webcams of the validation and test sequences** were not used during the **training phase**.

To evaluate the correspondance with the instrumental labels, we will use the Resnet50 pretrained on Imagenet with $p_o = 0.7$, $p_q = 0.3$ and a =0.1 (hinge margin). Its output is now referred to as our snow depth index. As we still meet problems with nighttime images, the following results only holds for the **daytime** part of the TENEBRE-q dataset.

Evaluation on TENEBRE-q (daytime images)

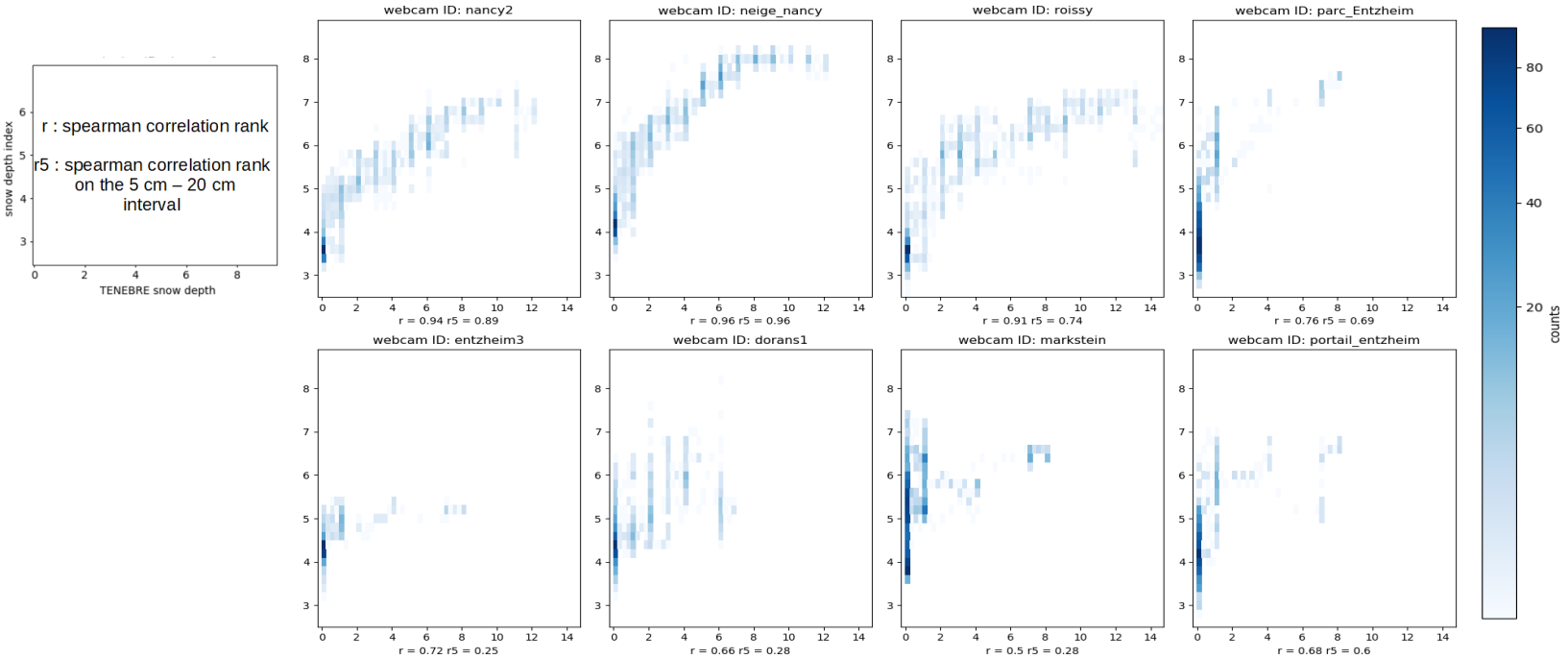


Fig.11. 2D histograms (TENEBRE-q snow depth vs. resnet50 outputs)

Two dimensional histograms show the extent to which our snow depth index co-vary with the instrumental labels. On these examples, the two first behave correctly. For both Roissy and Dorans1 webcams, the dispersion is stronger, and for the four lasts, predictions is more hazardous. Whatever the reasons (changes in the webcam orientation, instrumental default, etc), the main point is that the range of our snow depth index seems to be relative to the view. We are still working on that issue.

Finally, the rank correlations are lower over the 5cm – 15 cm interval. We ignore if it comes from the scarcity of such snow depths in our training dataset.

Evaluation on TENEBRE-q (daytime images) :

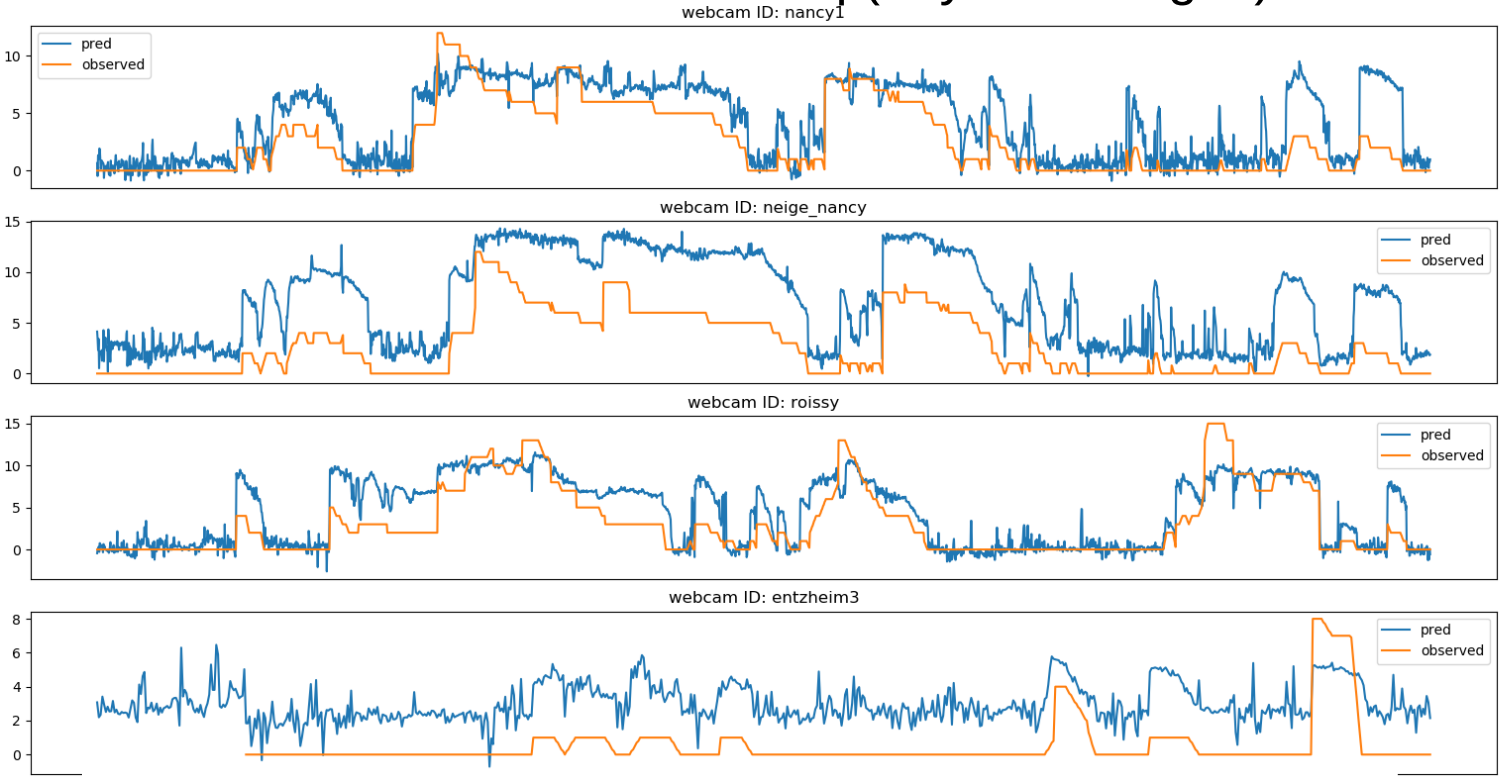


Fig.12. Coevolution of our snow depth index (after rescaling) and the instrumental labels. Strong discontinuities are partly due to concatenation of daytimes. The same affine rescaling has been used for all the curves.

Time series allow to observ the same phenomenon: if snow periods are well delimited by the index, the index range varies from one webcam to the other. Moreover, the prediction is affected by a high-frequency noise which may be inherent to our methodology or due to quick variations of correlated environmental factors, as the illumination.

Conclusion and perspectives

- We presented an end-to-end learning to rank framework that yield a snow depth index. On an independant subset of webcams located in weather stations, this index varies coherently with instrumental measures, at least by day, and for the first centimeters. The same approach gives similar results on the problem of the optical range estimation.

- However, the range of the output index varies from one webcam to another and from daytime to nighttime (not shown). It limits the interest for an application in road meteorology. Until now, our naive attempts to combine regression and ranking, in order to calibrate the estimations and improve the ranking correlation, led to overfitting. Nevertheless, customization of the training parameters could allow to benefit from both kinds of label, as suggested in [21].

- Finally, the time series of our snow index display a high-frequency noise that may trigger false detections. We ignore if this noise is inherent to our ranking method or if a better disentangling of correlated factors would help. In the second case, an extension to a semi-supervised framework with a constraint on time-variations may help (see [16]).

Bibliography

[1] V.Jacq, S.Balaguer. Neige ne plaine sur la Ferance méditerranéenne. Phénomènes remarquable n°10. Publication Météo-France.

[2] N. Jacobs, W. Burgin, N. Fridrich, A. Abrams, K. Miskell, B. H. Braswell, A. D. Richardson, and R. Pless, "The global network of outdoor webcams: Properties and applications" in Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ser. GIS '09. New York, NY, USA: ACM, 2009, pp. 111–120.

[3] M. Kosmala, K. Hufkens, and A. D. Richardson, "Integrating camera imagery, crowdsourcing, and deep learning to improve high-frequency automated monitoring of snow at continental-to-global scales" PLOS ONE, vol. 13, no. 12, pp. 1–19, 12 2018.

[4] Wang, M. Korayem, S. Blanco, and D. J. Crandall, "Tracking natural events through social media and computer vision" in Proceedings of the 24th ACM international conference on Multimedia. ACM, 2016, pp. 1097–1101.

[5] W.-T. Chu, X.-Y. Zheng, and D.-S. Ding, "Camera as weather sensor: Estimating weather information from single images" journal of Visual Communication and Image Representation, vol. 46, pp. 233 – 249, 2017.

[6] Elhoseiny, M., Huang, S., & Elgammal, A. (2015, September). Weather classification with deep convolutional neural networks. In 2015 IEEE International Conference on Image Processing (ICIP) (pp. 3349-3353). IEEE.

[7] W. Chu, K. Ho, and A. Borji, "Visual weather temperature prediction" CoRR, vol. abs/1801.08267, 2018

[8] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere", International Journal of Computer Vision, vol. 48, no. 3, p. 233–254,2002

[9] H.-Y. Zhou, B.-B. Gao, and J. Wu, "Sunrise or sunset: Selective comparison learning for subtle attribute recognition." CoRR, vol. Abs/1707.06335, 2017.

[10] Y. You, C. Lu, W. Wang, and C.-K. Tang, "Relative cnn-rnn: Learning relative atmospheric visibility from images" IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 45–55, 2019.

[11] C. Lu, D. Lin, J. Jia, and C.-K. Tang, "Two-class weather classification" in The IEEE Conference on Computer Vision and Pattern Recognition

[12] D. Glasner, P. Fua, T. Zickler, and L. Zelnik-Manor, "Hot or not: Exploring correlations between appearance and temperature",in 2015 IEEE International Conference on Computer Vision (ICCV), vol. 00, Dec. 2015, pp. 3997–4005.

[13]D. Lin, C. Lu, H. Huang, and J. Jia, "Rscm: Region selection and concurrency model for multi-class weather recognition", IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4154–4167, Sep. 2017.

[14] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[16] N. Graves and S. Newsam, "Camera-based visibility estimation: Incorporating multiple regions and unlabeled observations", Ecological informatics, vol. 23, pp. 62–68, 2014.

[17] Volokitin, A., Timofte, R., & Van Gool, L. (2016). Deep features or not: Temperature and time prediction in outdoor scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 63-71).

[20] Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS)

[21] Sculley, D. (2010, July). Combined regression and ranking. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 979-988).

[22] Wang, X., He, K., & Gupta, A. (2017). Transitive invariance for self-supervised visual representation learning. In Proceedings of the IEEE international conference on computer vision (pp. 1329-1338).

[23] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.