EGU Abstract

Accepted Paper

Department of Meteorology and Geophysics

Utilizing a residual neural network ensemble for ground-based cloud classifications

Markus Rosenberger[1], Manfred Dorninger[1], Martin Weissmann[1]

1) Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

1. Motivation and research questions

Human cloud observation coverage is stagnating or not available at all in many parts of the world. Since automated sky cameras have become inexpensive and widely available, automated cloud classification methods offer a unique solution to enable more consistent and homogeneous results around the world. Following research questions are addressed:

1. How accurately can a neural network ensemble retrieve cloud classes from ground based RGB images to compensate for gaps in human cloud observation coverage?
2. How can we overcome biases due to observation imbalances?
3. How reliable are the predicted probabilities of cloud class occurrence?

2. Data / Method

Data

- Around 12 000 instances with 4 pictures each → whole sky covered
- Class specific data augmentation to reduce observation imbalances → in total more than 20 000 instances
- Operational human SYNOP cloud observations as ground truth
- Up to 3 cloud classes per instance → multi-label classification

Method

- 10-member ensemble trained from scratch with identical architectures similar to ResNet[2] (cf. Fig. 1)
- Brier Score as loss function
- Random permutation of sub-images in each training epoch

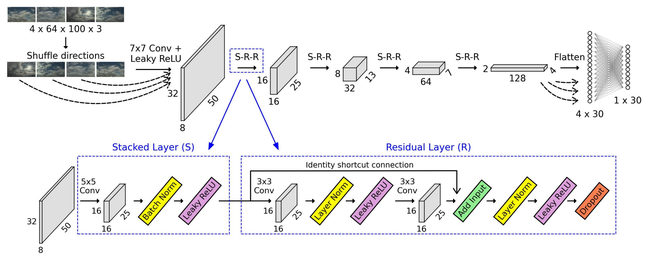
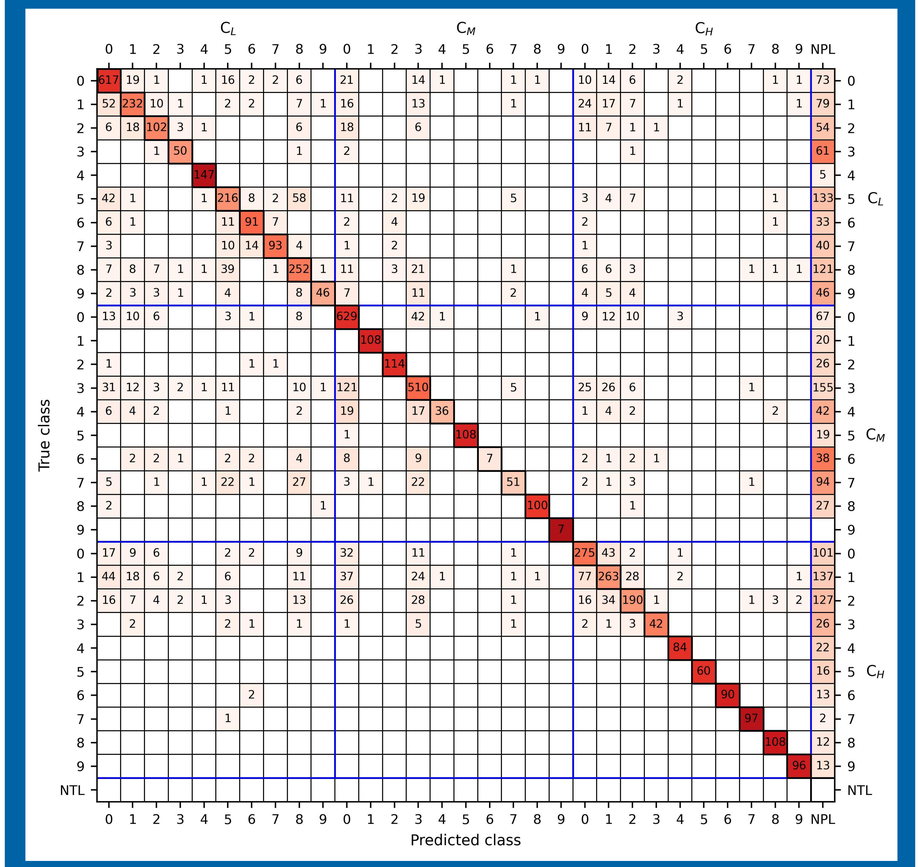


Figure 1. Sketch of the model architecture used for each ensemble member. Sub-images are processed individually, and respective outputs are combined in the final sigmoid layer.

3. Results

- Confusion matrix[3] of ensemble mean predictions indicates that True Positives dominate by far in majority of classes
- Data augmentation is crucial to get sufficient results but leads to reduced generalization ability in aggressively augmented classes
- Sub-image shuffling substantially enhances model robustness
- Predictions with too small probabilities (column *NPL*) represent the largest error source
- Worse performance in similar classes (e.g. C_M = 7 and C_L = 5) as well as in cases where temporal evolution is important (e.g. C_M = 6)

Ensemble mean can reliably & accurately classify clouds from RGB images into 30 SYNOP classes



The confusion matrix[3] of ensemble mean predictions indicates sufficient classification accuracy for most cloud classes. Column *NPL* contains predictions with too small predicted probabilities, making up the majority of false classifications.

3. Results continued

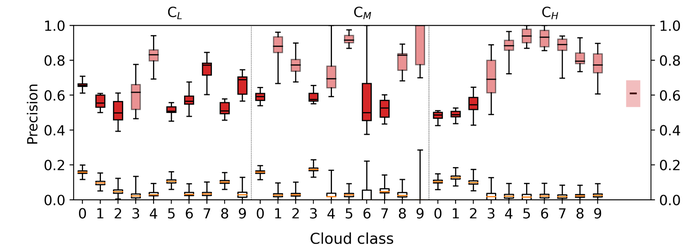


Figure 2. Distributions of precision scores of ensemble members (red boxes) and of 10 000 random forecasts (empty boxes). The patch on the right shows the distribution of weighted macro averages.

- Distributions of Precision, Recall, and MCC are well outside random prediction samples for each class (cf. Fig.2)
- Highest values in augmented classes (light shading in Fig. 2)
- Reliability diagrams indicate excellent reliability and resolution but underconfidence in least abundant classes

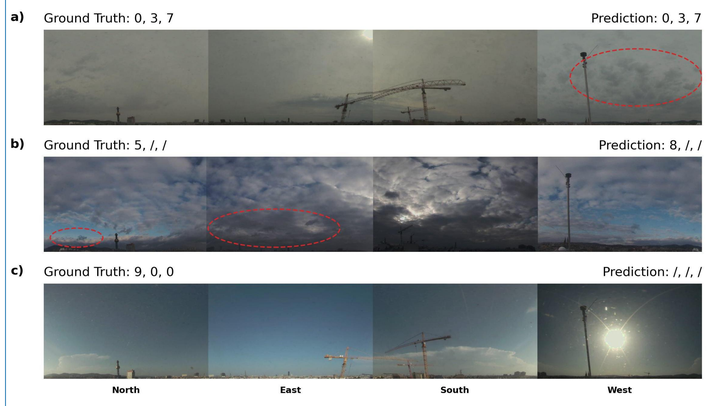


Figure 3. Example instances, where the prediction of our model (a) is perfectly accurate, (b) differs from the ground truth but fits the visible classes, and (c) lacks the correct class although it is easily visible.

4. Conclusions / Take home messages

1. Ensemble mean predicts cloud classes with high reliability and resolution and reaches a Precision of 0.83.
2. Each ensemble member outperforms both random and climatological predictions
3. Class specific data augmentation is crucial to reduce influence of observation imbalances, but it can lead to overfitting
4. Aggressively augmented classes show highest scores but also underconfidence in reliability diagrams

5. References

[1] Rosenberger, M., Dorninger, M., & Weissmann, M. (2025). Deriving WMO cloud classes from ground-based RGB pictures with a residual neural network ensemble. *Earth and Space Science*, 12, e2024EA004112, accepted.

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv.

[3] Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, *10*, 19083–19095.