markus.rosenberger@univie.ac.at

universität wien

Department of Meteorology and Geophysics

Using cGAN for cloud classification from RGB pictures

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

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

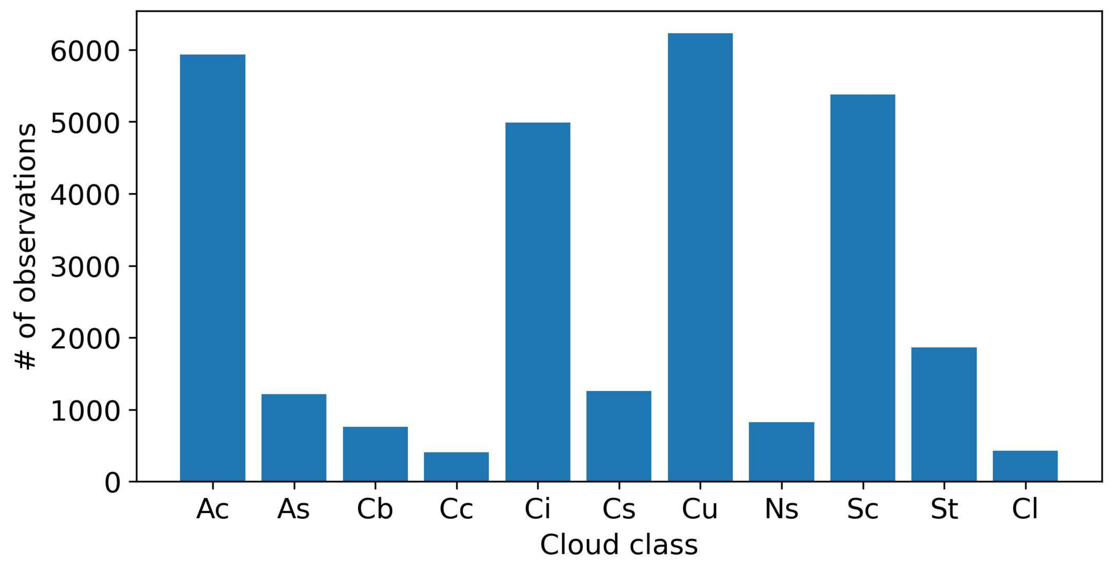
1. Motivation and research questions

The number of human cloud observers is decreasing and many parts of the world are not covered by cloud observers at all. Automated sky cameras, however, have become inexpensive and widely available. Hence, automated cloud classification methods would be easy to apply around the world and deliver homogeneous results.

1. Can machine learning algorithms retrieve cloud classes from RGB pictures in a reliable way?
2. How to deal with biases because of highly imbalanced classes?

2. Methods & Data

- Convolutional neural networks have proved to deliver sufficient results in image classification tasks
- Conditional Generative Adversarial Networks (cGANs, Goodfellow et al., 2014) can create artificial images belonging to specific classes and thus reduce the problem of class imbalances
- Panorama images are taken with department's cloud observation system from 05.10.2016 – 18.02.2019 and from 04.05.2022 onwards
- Ground truth observations are hourly operational observations at station Vienna Hohe Warte
- Conversion from 30 observed SYNOP classes to 10 cloud genera + clear sky (Cl) → multi-class multi-label classification problem
- Asymmetric Loss (ASL; Ben-Baruch et al. 2021) for multi-label classification, binary cross-entropy for real/fake discrimination
- ASL weighted to account for class imbalances in data set
- Parameter introduced to decide how much fake label information is used for discriminator training



3. Results

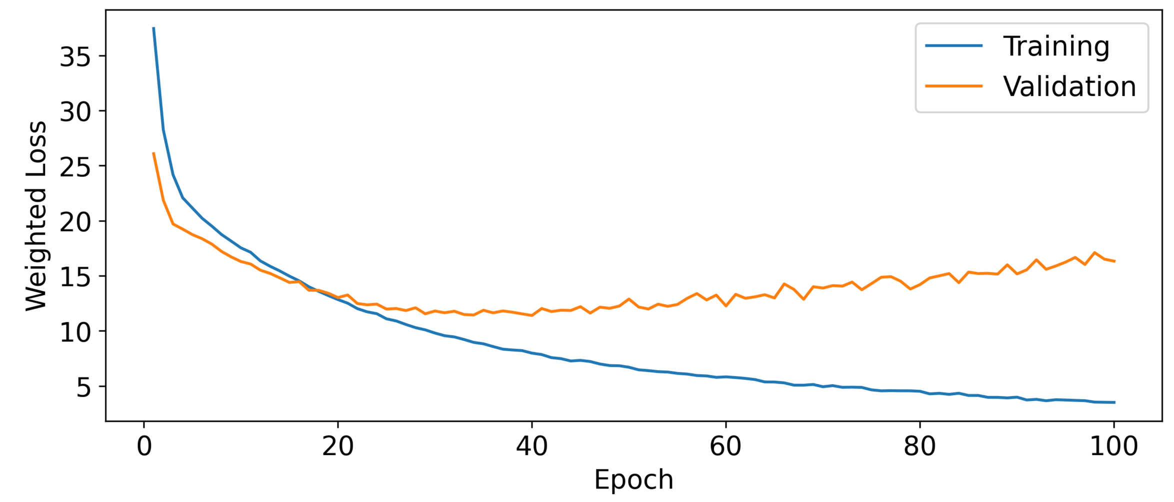
- Multi-label confusion matrix according to Heydarian et al. (2022) to investigate performance on validation data
- True positives dominate by far in almost all classes
- Highest false positive rates in most abundant classes (e.g. Ac, Ci)
- Highest false negative rates in classes with visual similarities (e.g. Cb vs. Cu) or classes with joint occurrences (e.g. Sc vs. Ac and Cu)
- Generator architecture needs to be improved to create more realistic images
- Discriminator architecture needs fine tuning for higher precision

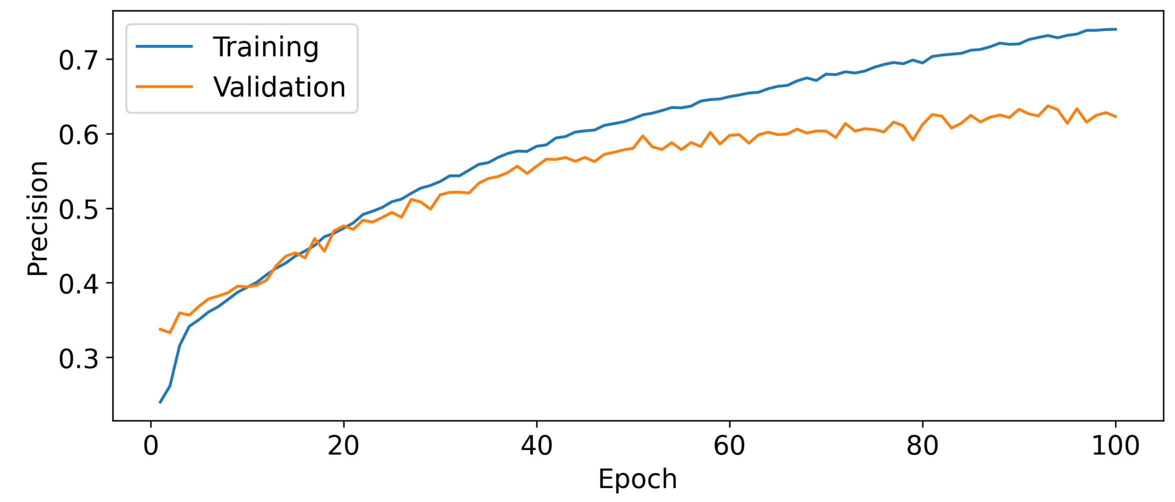
A cGAN can be trained to classify clouds on RGB images with high precision and recall

True class \ Predicted class	Ac	As	Cb	Cc	Ci	Cs	Cu	Ns	Sc	St	Cl	NPL
Ac	53	1	2		11	2	13		10	3	2	2
As	4	79	2		2		7		3	3		1
Cb	14	1	25		10	4	29		15	1		
Cc	3			75	8	3	8		2			
Ci	12		3		52	2	13		9	1	4	2
Cs	2		1		2	88	4		1	2		
Cu	13	1	5		13	2	53	1	5	2	1	2
Ns	2		1		1		9	79	4	4		
Sc	14	2	2		11	2	14	2	44	7	1	2
St	4	4			1		15	3	8	63		
Cl	7		1		31		8		8	4	41	
NTL												

The multi-label confusion matrix indicates that a cGAN can predict the occurring cloud classes from RGB pictures in most cases with a sufficient precision and recall. Inaccuracies occur because of highly imbalanced ground truth classes (Ac, Ci, Cu, Sc) as well as similar visual properties of different classes (e.g. Cb and Cu).

3. Results continued





- Training for more epochs would further increase precision on the training data but overfitting already starts to be a serious issue
- Generated images still need a lot of improvement to be used for data set enlargement

Upper panel: Image taken from the training data set. **Lower panel:** Image created by the generator after 100 training epochs.

4. Conclusions

1. True positive rate already high after a few training epochs
2. Biases due to class imbalances are reduced but still present
3. Visual similarities and joint occurrence of classes on images introduce inaccuracies (false negatives & false positives)
4. Generator architecture needs large improvements
5. Discriminator has to be regularized even more

5. References

[1] Ben-Baruch, E., Ridnik, T., Zamir, N., Noy, A., Friedman, I., Protter, M., and Zelnik-Manor, L. (2021). Asymmetric loss for multi-label classification.

[2] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks.

[3] Heydarian, M., Doyle, T. E., and Samavi, R. (2022). Mlcm: Multi-label confusion matrix. IEEE Access, 10:19083–19095.

Using cGAN for cloud classification from RGB pictures – Supplementary material

Markus Rosenberger - markus.rosenberger@univie.ac.at

First of all: If you have any questions, comments or even suggestions regarding our work, please do not hesitate to contact us via the above given address. We are happy about any contributions.

Different machine learning methods have already been used to automatically retrieve cloud types from satellite data during the past decades. However, since the WMO defined cloud observation standards in their cloud atlas via visual properties seen from the Earth's surface, the results of these previously mentioned methods can hardly be compared directly to human observations. Thus in this work we try to find a way to automatically retrieve cloud classes from conventional RGB pictures taken at the Earth's surface. Convolutional neural networks (CNNs) proved to deliver sufficient results in diverse image classification tasks. And since there is a large imbalance in the occurrence of different cloud classes, e.g. cumulus clouds are observed much more frequently than nimbostratus, a method has to be found that can compensate this imbalance.

Generative Adversarial Networks (GANs; Goodfellow et al., 2014) are a combination of 2 CNNs competing against each other during the training process with the ultimate goal that one of the networks (the generator G) can artificially create images within the distribution of the training data. The only input needed is a seed of random numbers. A special architecture called conditional GANs (cGANs) is even able to create fake images corresponding to a specific class of the data set. In this case, in addition to the noise seed the generator has to be provided a ground truth label. The aim of the other network (the discriminator D) is to decide if its input image is either drawn from the training data or generated by G. In addition to the real/fake decision, the discriminator of a cGAN can also be trained to classify the images it is fed with. During one step of the training, G is only trained on how well it was able to trick D into believing the generated images were drawn from the training data set. On the other hand, D is trained on both its ability to discriminate real from fake pictures but also on its classification score on both real and fake pictures. However, since at the beginning of the training process the fake images just consist of random noise, it may happen that the D learns the labels of the fake images faster than those of the real ones. Ultimately this could lead to the result that the classification skill for the training set is close to zero and no information can be gained. Hence, we introduce a hyperparameter α that controls how much of the fake label information is used to train the discriminator at each epoch.

Another way to account for class imbalances is a weighted loss function. Ben-Baruch et al. (2021) introduced the Asymmetric Loss (ASL) which takes into account that in most multi-label classification scenarios less labels are actually observed in an instance than not. Moreover, the predicted probability of each class is taken into account in every instance to differently weigh, e.g. strong and weak misclassifications. The loss value calculated using ASL is in our work additionally weighted by the relative abundance of the observed classes in each instance.

The aim of this work is to train a cGAN to:

1. Discriminate between 11 cloud classes with high precision and reliability
2. Generate images of specific cloud classes with a quality high enough that they can be added to the training data set

To achieve these goals a proper data set is needed. We use cloud images from the department's cloud observing system (Icos) which consists of 4 cameras pointing in the main cardinal directions. Images are available in 2 distinct periods, from 05.10.2016 — 18.02.2019 and from 04.05.2022 onwards and are used as panoramas during training and evaluation of our models. Hourly operational cloud observations at the station Vienna Hohe Warte are used as ground truth since this station and the cameras are less than 2 km apart. The main advantage of taking images from a single camera system rather than from several different sources, is that the model benefits from a homogeneous data set.

In order to evaluate the performance of the discriminator on the classification task, the Multi-Label Confusion Matrix (MLCM; Heydarian et al., 2022) is used. The MLCM is evaluated on the validation data set and true positive (TP) classifications are located on the main diagonal. Apart from the main diagonal false negative (FN) rates for each observed class are shown in the corresponding row and false positive (FP) rates are shown in the columns of the matrix.

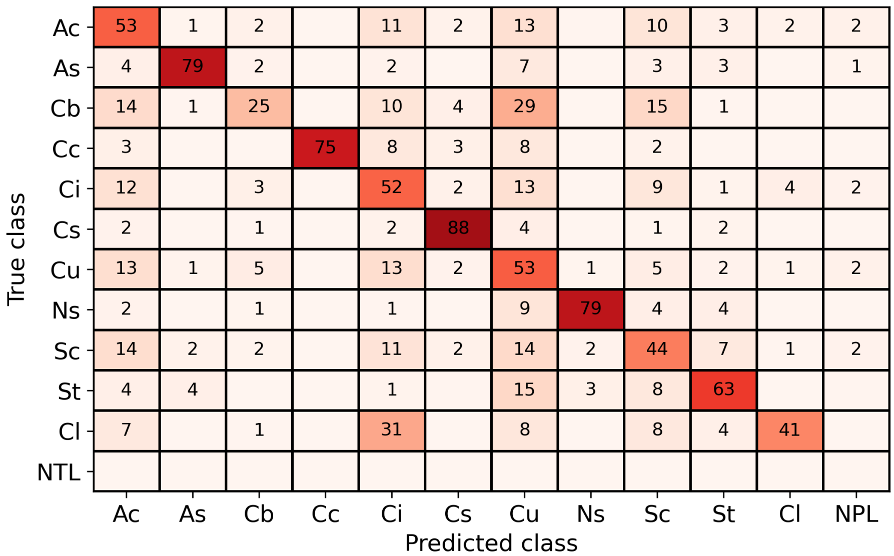


Figure 1: MLCM indicating that in many classes recall values are very high. Inaccuracies can be found either in classes with a very large number of observations (e.g. Ac, Cu), in classes that tend to occur together (Sc together with Ac or Cu) or in classes with visual similarities (e.g. Cb and Cu).

In 8 out of 11 classes, recall is above 50% and in 4 of them it is even $\geq 75\%$. On the other hand, there is only one class with a very small recall: Cb with a value of 25%. In this case, the FN rate of a predicted Cu cloud where Cb is observed is even higher than the recall rate. This may have two different reasons: Firstly, Cb is one of the least abundant classes, which is supposed to lead to a general bias in the classification. Secondly, Cu and Cb have several visual similarities like the vertical development which often resembles a cauliflower. One way of differing between these two genera is by the absence of lightning and thunder when Cu are observed, which is, however, a property the algorithm cannot detect solely from the images. Thus, the confusion is understandable. Those classes that have a high false positive rate (e.g. Ac and Cu) are again the same that are observed very often. Summed up, one can say that the discriminator's performance is already very good on many different classes but that class imbalances are still an issue, though probably to a less extent than without the weighted loss function.

Regarding the artificially created images, currently there is no sufficient method to objectively evaluate the performance of the generator. Especially for the case where the images shall be used as additional input for a classification model, the quality and reliability should be as high as possible. Finding a threshold in any metric that defines the conversion from *Not good enough* to *Good enough* for the inclusion in the data set will be a future task. However, at the current state of our project, the generator is by far not able to fulfill this task, which can be easily seen without any metric. On the one hand, not many structures are visible in the images and on the other hand there is a spurious grid over the whole picture domain, whose origin we do not know yet. Hence, the next step will be to improve the generator's architecture and thus also to improve the image quality.

To conclude, we have shown that in principal a cGAN can be extended in order to be used as a classifier for RGB cloud images with a very high precision and recall in many classes, although there are still some classes with significantly smaller values. The trend of the loss function during the training process indicates that there is also the need for regularization of the discriminator in order to further improve its performance. In addition, there is still a lot of work to do concerning the generator so that it is able to create usable pictures.

References

Ben-Baruch, E., Ridnik, T., Zamir, N., Noy, A., Friedman, I., Protter, M., and Zelnik-Manor, L. (2021). Asymmetric loss for multi-label classification.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks.

Heydarian, M., Doyle, T. E., and Samavi, R. (2022). Mlcm: Multi-label confusion matrix. *IEEE Access*, 10:19083–19095.