A Self-supervised Approach for Cloud Image Analysis*

Bhupendra A. Raut†, Dario Dematties‡, Robert C. Jackson, Seongha Park, Sean Shahkarami, Yongho Kim, Rajesh Sankaran, Pete Beckman, Scott M. Collis, Nicola Ferrier

Northwestern Argonne Institute of Science and Engineering, Northwestern University, Evanston, 60208, IL, USA

Argonne National Laboratory, Lemont, IL 60439, USA

January 6, 2023

Abstract

Clouds play a significant role in the global climate system, influencing the energy balance of the Earth's atmosphere through absorption, reflection, and scattering of incoming shortwave solar radiation, and absorption and emission of outgoing longwave radiation. Different types of clouds result from different atmospheric processes, and accurately identifying and analyzing cloud coverage is essential for understanding the Earth's radiative budget. While traditional computer vision methods and newer deep neural network (DNN) approaches have been used for cloud segmentation and categorization, these methods often rely on subjective human labeling of cloud or clear sky pixels for training. This can bias the algorithm's performance or hinder the extraction of relevant information from sky images. In addition, labeling cloud images can be a tedious process. In this work, we present a self-supervised learning approach for autonomously extracting relevant features from sky images captured by ground-based cameras for the classification and segmentation of clouds. Our approach aims to overcome the limitations of traditional methods and DNNs by using self-knowledge distillation and regularization to learn features from the data without the need for human labeling. We evaluate the performance of our approach on two datasets and demonstrate its ability to accurately classify and segment clouds.

Extended Abstract

In machine learning, two popular strategies are supervised and unsupervised learning. Supervised learning involves adjusting the parameters of the algorithm by comparing the predictions made by the algorithm to ground truths, or actual labels provided by humans for each sample in the data. Unsupervised learning does not have labels, only samples. The algorithms must therefore adjust their parameters using the statistical distribution of the data itself. Self-supervised learning is a combination of supervised and unsupervised learning. It is similar to unsupervised learning in that it does not require labels provided by humans in the data set. However, it is similar to supervised learning in that it uses labels, but they are self-extracted from the data set rather than provided by humans. Our self-supervised learning approach utilizes a joint embedding architecture and self-knowledge distillation to learn features from sky camera images without the need for human labeling (Caron et al., 2020, 2021; Dematties et al., 2022).

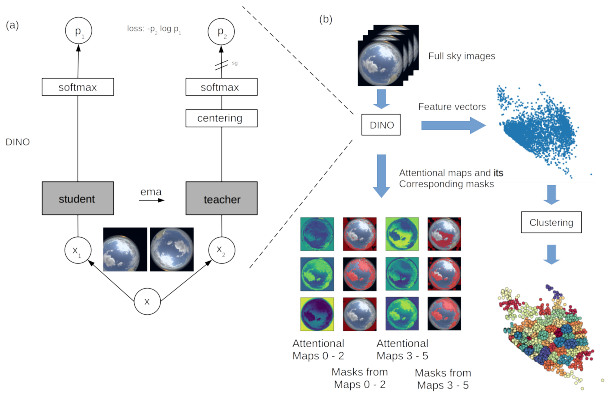


Figure 1: A joint embedding model architecture.

This architecture consists of two branches, each containing a network (Fig. 1). The networks in the branches can share weights, have different weights but the same architecture, or have completely different architectures. Our model consists of two Visual Transformers (ViT) using the same architecture. The transformers allows for the creation of attentional maps, which can be used to generate semantic segmentation of images without labels. One network is called the student and the other is the teacher. Both networks are randomly initialized, but only the student backpropagates errors in an attempt to mimic the output of the teacher. The student tries to follow the frozen teacher, adjusting

*22nd Conference on Artificial Intelligence for Environmental Science in 103rd AMS Annual Meeting, 08–12 January 2023, Denver CO. https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/413343

†Presenting author: B.A. Raut, braut@anl.gov

‡Corresponding author: D. Dematties, ddematties@anl.gov

its weights progressively, while the teacher makes inferences using self-generated weights. Occasionally, the teacher updates its weights using the student weights, incorporating them into its architecture using exponential moving average (EMA).

The goal of the task, as determined by the loss function, is to produce outputs as similar as possible in both branches, regardless of the augmentations present in the input. The idea is that the network learns to ignore changes in color, brightness, size, position, etc. and instead focuses on the main features of the original image to invariantly recognize the objects present in it. However, the SSL approach can sometimes result in collapse, where the network ignores its inputs and produces identical constant outputs in both branches. To avoid collapse, the model uses stop gradient, centering, and sharpening, as well as EMA. Attentional maps, which are created using transformers, can be used to conduct both quantitative and qualitative analysis of the network's behavior. By examining the attentional maps, masks can be generated for clouds, sun, and clear sky. The different clouds may be tracked in the different attention maps using cloud tracking methods (Raut et al., 2022). We apply a two-dimensional self-organizing map grid of 5 by 4 nodes on this new feature space and classified each image in the dataset using the k-nearest neighbor algorithm based on the prediction inference from the self-organizing map nodes. Our approach demonstrates the ability to discriminate full-sky images based on cloud coverage, diurnal variation, cloud base height, and semantically segment cloud areas without labels (Dematties et al., 2022). We test the approach on two datasets: one with approximately 85,000 images collected from our ground-based camera and another with 400 labeled images from the WSISEG-Database following (CV-Application, 2021; Park et al., 2021). We also show examples of attention maps for satellite images (Fig. 2).

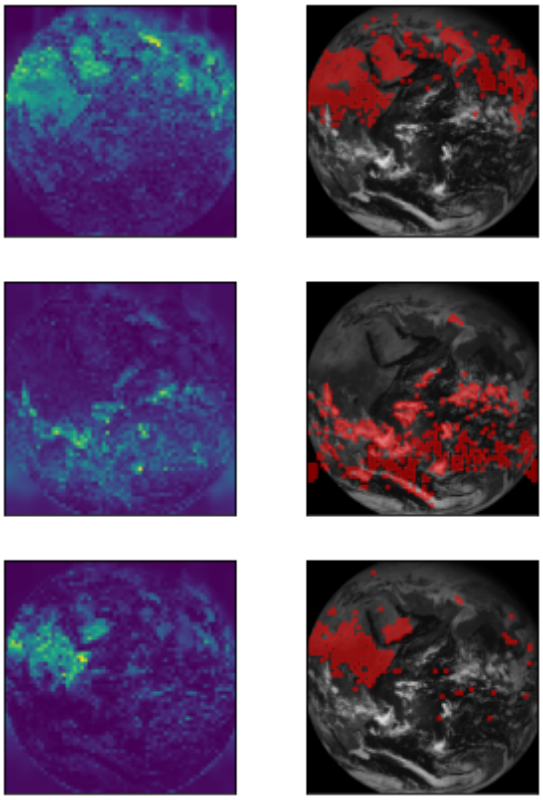


Figure 2: Attention maps showing different regions in the Meteosat-5 VIS image.

Acknowledgments

The Sage project is funded through the U.S. National Science Foundation's Mid-Scale Research Infrastructure program, NSF-OAC-1935984. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. The code to reproduce the experiments presented in this manuscript can be found at: `https://github.com/dariodematties/cloud_dino`.

References

Caron, M., I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, 2020: Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems*, **33**, 9912–9924.

Caron, M., H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, 2021: Emerging properties in self-supervised vision transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9650–9660.

CV-Application, 2021: WSISEG-Database. URL https://github.com/CV-Application/WSISEG-Database, original-date: 2019-01-24T08:00:05Z.

Dematties, D., and Coauthors, 2022: Let's unleash the network judgement: A self-supervised approach for cloud image analysis. *Artificial Intelligence for the Earth Systems*, **00**, In 2nd review.

Park, S., Y. Kim, N. J. Ferrier, S. M. Collis, R. Sankaran, and P. H. Beckman, 2021: Prediction of solar irradiance and photovoltaic solar energy product based on cloud coverage estimation using machine learning methods. *Atmosphere*, **12 (3)**, 395.

Raut, B. A., and Coauthors, 2022: Phase correlation on the edge for estimating cloud motion. *Atmos. Meas. Tech. Discuss. [preprint in review]*, 1–18, https://doi.org/10.5194/amt-2022-159.