Self-supervised detection of atmospheric phenomena from remotely sensed synthetic aperture radar imagery

Yannik Glaser
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu, HI 96822
yglaser@hawaii.edu

Peter Sadowski
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu, HI 96822
peter.sadowski@hawaii.edu

Justin E. Stopa
Ocean Engineering Department
University of Hawai'i Mānoa
stopa@hawaii.edu

Abstract

The European Space Agency provides unprecedented monitoring of Earth's oceans through a network of Synthetic Aperture Radar (SAR) satellites called Sentinel-1. Imagery from these satellites captures a variety of atmosphere and ocean surface phenomena including waves, atmospheric turbulence, ocean fronts, and marine biology. Computer vision methods have been used to process the large number of acquired images, but the use of machine learning methods has been severely limited by sparsely labeled data. Consequently, we apply a self-supervised learning method, SwAV, to three years of Sentinel-1 satellite observations (3 million images) to learn an unsupervised embedding for SAR images, then fine-tune the model to detect wind streaks and mesoscale convective cells through supervised learning. Our results demonstrate detection performance improvement over the previous state-of-the-art model but suggest that self-supervised training has marginal improvements over a more standard approach of transfer learning from a model trained on natural images.

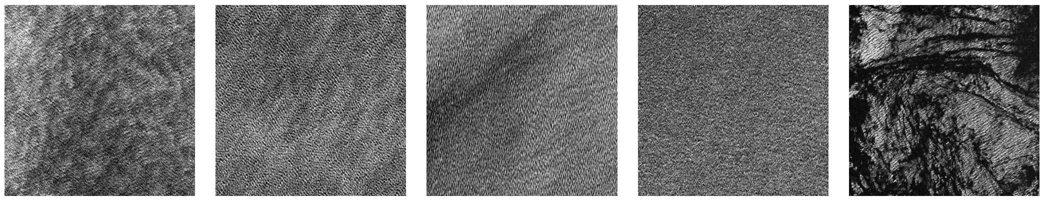
1 Introduction

Remote sensing missions such as the Sentinel-1 mission operated by ESA through the Copernicus Programme [1] produce huge amounts of satellite imagery data at unprecedented resolution and coverage. This data is valuable for a variety of scientific and environmental conservation purposes, such as land motion, polar science, and tropical cyclone monitoring, but fully leveraging the data requires automated analysis tools. Computer vision systems based on deep learning have proven tremendously useful for this task, but the reliance on labeled training data remains a bottleneck. In many scientific domains, labeling can only be performed by trained experts, so labeling more than a few thousand images is infeasible.

Self-supervised, contrastive representation learning is a promising method for alleviating this bottleneck. In this approach, a deep neural network is trained on large amounts of *unlabeled* data to learn a representation (embedding) that is useful for downstream tasks. While contrastive learning has been shown to produce useful embeddings for natural images [10, 16, 9] and medical images [6], it has not been widely applied to remote sensing data. This study tests the hypothesis that contrastive learning can help overcome the challenges of limited labeled training data for SAR image analysis and is expected to be relevant for other remote sensing datasets.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Here, we focus on data collected by the Sentinel-1 (S-1) mission, a part of the Copernicus Programme, which launched two satellites, S-1 A and B, in April 2014 and 2016 respectively [23]. Sentinel-1B ceased operation in December 2021 and Sentinel-1C is planned to be launched in early 2023. These satellites are equipped with a C-band synthetic aperture radar instrument to perform continuous high-resolution day-and-night imaging unaffected by cloud cover. One of the four acquisition modes of the SAR sensor, primarily for measuring ocean waves across the global open ocean, is the 'WaVe' mode (WV). Examples of WV images are shown in Figure 1. Between the two S-1 satellites, nearly 120,000 WV images of the ocean surface are collected every month, each covering 20 km by 20 km at 5 m resolution.



(a) MC (b) MC/WS (c) WS (d) NV (e) OT

Figure 1: Example SAR images (20 × 20 km) showing the continuum of atmospheric stability in order of increasing stability: (a) strongly unstable - micro-scale convective cells (MC); (b) unstable - mixed convection cells and wind streaks (MC/WS); (c) near-neutral - wind streaks (WS); (d) stable - negligible atmospheric variability (NV). The last image (e) shows an example of a biological surface slick, which is labeled as "other" in the present study.

The S-1 WV capture a variety of geophysical phenomena beyond ocean surface waves [4, 11], including atmospheric phenomena like mesoscale convection cells (MC) [5], wind streaks (WS) [27, 14, 19, 18], and many others [29, 3, 13, 17]. Stopa et al. [22] have shown that MC, WS, and negligible atmospheric variability (NV) represent a spectrum of stability regimes along the oceanic surface layer, corresponding to unstable, near stable, and stable conditions respectively. These conditions correspond to the atmospheric stratification in response to both thermal and shear forcing [12], and are thus relevant to a number of applications such as air-sea fluxes studies, planetary boundary layer dynamics, wind energy resource assessment, and relationship with cloud dynamics essential in climate modeling.

Previous work by Stopa et al. [22] classified SAR images into these three categories of ocean surface stability using the convolutional neural network model from Wang et al. [25] (CmWV). The training dataset is thought to be biased, with only exemplary images selected for each class. In addition, the protocol only allowed for a single class label per image, even though multiple phenomena typically impact the sea surface roughness measured by SAR and present in the images. This resulted in a non-representative dataset that did not reflect the true WV population.

This study addresses those issues with a new hand-labeled dataset of multi-output classification labels, and analyzes self-supervised contrastive learning as an approach for handling the problem of limited labels. A dataset of 2,200 unbiased, hand-labeled, multi-label observations was compiled by several experts. We then leveraged three years' worth (3 million images) of unlabeled S-1A and S-1B WV images to learn a SAR image embedding model with contrastive learning, and fine-tuned the learned embedding model to classify sea surface stability. This approach is compared to the CmWV model and a standard transfer learning approach of fine-tuning a ResNet-50 model [15] pre-trained on natural images.

2 Methods

2.1 Contrastive Learning Framework

Contrastive self-supervised representation learning has seen a recent surge in popularity, especially for computer vision tasks, because of the impressive performance of the learned representations on downstream tasks and the ability to leverage large amounts of unlabeled data (e.g. Chen et al. [10], Caron et al. [8]). At a high level, most modern contrastive learning methods are based on objectives that group representations of multiple views (obtained by applying different augmentations) of the same sample closely together while representations of other examples in the dataset are pushed away. Sometimes other so-called pretext tasks may also be used, such as the jigsaw task proposed by Noroozi and Favaro [20].

Swapping Assignments between multiple Views (SwAV) is a contrastive framework proposed by Caron et al. [8] with the desirable quality of not requiring prohibitively large batch sizes, a common limitation exhibited by frameworks such as the popular SimCLR [10]. While other solutions have been proposed to this problem, such as momentum encoders [16], these introduce high complexity and memory requirements during training. SwAV is based on a clustering pretext task, where representations are assigned to clusters and the network has to learn to predict cluster assignments. This framework works with small batch sizes by storing cluster assignments from past batches in a queue, enabling training without the need for large compute clusters.

For this work, we adopt the SwAV framework with the parameterization proposed in the original paper, using a batch size of 1024 samples distributed over eight NVidia V100 32GB GPUs, a queue length of 16 batches, and 1000 cluster centroids. A standard ResNet-50 [15] backbone is used. Training was stopped after 65 epochs, 10 days, due to time and computational constraints.

To evaluate representation quality on the downstream task, we compared three common supervised learning methods. First, classification was performed using a weighted k-nearest neighbor (kNN) classifier with the Euclidean distance between the embeddings, as in Caron et al. [9] and Wu et al. [28]. This method is robust to hyperparameter choices as there is only one parameter (k) that can be chosen based on an exhaustive evaluation of a validation set, as such it serves as a simple and effective quality measure for representations. Second, we used the linear evaluation protocol from Chen et al. [10], which locks all weights in the trained ResNet and trains a single softmax layer to perform classification from the embeddings with stochastic gradient descent until no improvement in validation loss is observed for 10 epochs. This was performed using both the ImageNet and the SwAV weights for comparison. Lastly, a similar setup as in the previous protocol is used but all weights in the network are fine-tuned on the labeled dataset. Fine-tuning is performed until no improvement in validation loss is observed for 50 epochs. For the last protocols, hyperparameters were chosen in accordance to the original paper [8]. Training is done on a single V100 GPU with a batch size of 128.

2.2 Datasets

Unlabeled dataset. The satellite imagery dataset consists of three years (2017, 2018, 2019) of Copernicus S-1 A and B observations. Retrieval and processing mostly followed the protocol described in Wang et al. [25] and Stopa et al. [22]. No further pre-processing is done except zero-padding all images to a uniform (450 by 450 pixels) size. Overall, this dataset contains 2,943,550 images, of which 90% is used for training and 10% is used for validation. All SAR WV image are obtained from the ESA's Sentinel Open Access Hub [2] where they are freely available without licensing restrictions.

Labeled dataset. The labeled data used here is designed to address several shortcomings of the original work by Wang et al. [25], specifically, it is multi-label, allowing for multiple positive classes per WV vignette, and sampled randomly to be more representative of the underlying distribution of atmospheric phenomena over the global ocean, as detailed in Stopa et al. [22]. Images were labeled by at least one expert, and a consensus label was produced for each image to serve as the ground-truth for this study. Phenomena besides WS, MC, and NV were also labeled for this dataset (e.g. ice bergs, ocean swells, etc.), but not used in the current study. For the surface layer stability estimation task, four classes of interest are used: (1) MC, (2) WS, (3) NV, and (4) other (OT). Since WS and MC exist on a continuum of surface layer stability, there are multi-label cases in which both classes are positive. The NV label is defined as the absence of WS, MC, and any other phenomenon. Lastly, OT is the absence of WS and MC, but the presence of another labeled phenomenon. In total, this dataset contains 2,300 samples which are stratified and split into 60% training data, 20% validation data for early stopping and hyperparameter tuning, and 20% held-out test data for final model evaluation per class.

3 Results

Preliminary results are summarized in Table 1. The contrastive learning method shows minor improvements for both evaluation settings where no further fine-tuning of the main ResNet weights is done. However, the best-performing models are the ones resulting from end-to-end fine-tuning of all weights in the network, where the ImageNet weights appear to actually perform marginally better than the contrastive weights, with the best micro-averaged area under the receiver operating characteristic (AUROC) being 0.93. However, all performance differences here are small. For the WS and MC classes, both of our models significantly outperform the CmWV model by Wang et al. [26] in all evaluation scenarios, for the NV class, all models perform about equally strong. Because labeling criteria and classes differed between the Wang et al. [25] dataset and the dataset used here, comparison for the OT class was not possible.

	Contrastive Weights	ImageNet Weights
kNN	**0.864**	0.859
Linear Evaluation	**0.841**	0.836
Finetuning	0.929	**0.931**

(a)

	MC	WS	NV	OT
SwAV Weights	0.872	0.831	0.952	**0.910**
ImageNet Weights	**0.873**	**0.850**	**0.960**	0.905
CmWV	0.793	0.727	0.946	-

(b)

Table 1: Performance summary. (a) Micro AUROCs for different evaluation stratgies. (b) Micro AUROCs for finetuned models on individual classes.

4 Discussion

This work presents an initial exploration of self-supervised contrastive learning for satellite data. A large dataset of unlabeled SAR WV observations is compiled for contrastive training and the resulting model is evaluated for the most commonly occurring sea surface roughness images related to the atmosphere above. Initial experiments do not support the use of contrastive learning, with improvements over transfer learning from natural image features being marginal or not present at all, while requiring significant computation to train. However, these results are preliminary, and longer training times, more extensive hyperparameter tuning, and selecting pretext-tasks more relevant to remote sensing data need to be explored before definitive conclusions can be reached. Semi-supervised approaches such as contrastive learning that leverage large amounts of unlabeled image data could be extremely valuable in remote sensing applications where annotated training data can only be produced by experts.

Our results, both with contrastive and ImageNet transfer learning, do show drastic detection performance improvement over previous models. This has immediate consequences for ocean surface layer estimation, where providing a highly reliable model can help with automated analysis and downstream remote sensing applications. We plan to release the model and code at the conclusion of this project. While advancing SAR analysis has many potentially large benefits, such as improving understanding of the planetary boundary layer which would ultimately improve climate models, or sea ice /oil spill monitoring, there is potential for maritime and ground surveillance [21, 24] and military applications [7] based on SAR data also improving as a consequence of advances in correctly classifying the SAR imagery.

References

[1] Copernicus. URL `https://www.esa.int/Applications/Observing_the_Earth/Copernicus`.

[2] Sentinel access hub. URL `https://sentinel.esa.int/web/sentinel/sentinel-data-access`.

[3] Werner Alpers, Biao Zhang, Alexis Mouche, Kan Zeng, and Pak Wai Chan. Rain footprints on c-band synthetic aperture radar images of the ocean-revisited. *Remote Sensing of Environment*, 187:169–185, 2016.

[4] Fabrice Ardhuin, Bertrand Chapron, and Fabrice Collard. Observation of swell dissipation across oceans. *Geophysical Research Letters*, 36(6), 2009.

[5] BW Atkinson and J Wu Zhang. Mesoscale shallow convection in the atmosphere. *Reviews of Geophysics*, 34(4):403–431, 1996.

[6] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3478–3488, 2021.

[7] Carole Belloni, Alessio Balleri, Nabil Aouf, Thomas Merlet, and Jean-Marc Le Caillec. Sar image dataset of military ground targets with multiple poses for atr. In *Target and background signatures III*, volume 10432, pages 218–225. SPIE, 2017.

[8] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments, 2021.

[9] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021.

[10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.

[11] Fabrice Collard, Fabrice Ardhuin, and Bertrand Chapron. Monitoring and analysis of ocean swell fields from space: New methods for routine observations. *Journal of Geophysical Research: Oceans*, 114(C7), 2009.

[12] James B Edson, Venkata Jampana, Robert A Weller, Sebastien P Bigorre, Albert J Plueddemann, Christopher W Fairall, Scott D Miller, Larry Mahrt, Dean Vickers, and Hans Hersbach. On the exchange of momentum over the open ocean. *Journal of Physical Oceanography*, 43(8): 1589–1610, 2013.

[13] Heidi A Espedal, Ola M Johannessen, and Johan Knulst. Satellite detection of natural films on the ocean surface. *Geophysical Research Letters*, 23(22):3151–3154, 1996.

[14] TW Gerling. Structure of the surface wind field from the seasat sar. *Journal of Geophysical Research: Oceans*, 91(C2):2308–2320, 1986.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning, 2020.

[17] T Jia, JJ Liang, X-M Li, and J Sha. Sar observation and numerical simulation of internal solitary wave refraction and reconnection behind the dongsha atoll. *Journal of Geophysical Research: Oceans*, 123(1):74–89, 2018.

[18] Wolfgang Koch. Directional analysis of sar images aiming at wind direction. *IEEE Transactions on Geoscience and Remote Sensing*, 42(4):702–710, 2004.

[19] Susanne Lehner, Jochen Horstmann, Wolfgang Koch, and Wolfgang Rosenthal. Mesoscale wind measurements using recalibrated ers sar images. *Journal of Geophysical Research: Oceans*, 103(C4):7847–7856, 1998.

[20] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision*, pages 69–84. Springer, 2016.

[21] Ramona Pelich, Nicolas Longépé, Grégoire Mercier, Guillaume Hajduch, and René Garello. Ais-based evaluation of target detectors and sar sensors characteristics for maritime surveillance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8): 3892–3901, 2014.

[22] Justin E Stopa, Chen Wang, Doug Vandemark, Ralph Foster, Alexis Mouche, and Bertrand Chapron. Automated global classification of surface layer stratification using high-resolution sea surface roughness measurements by satellite synthetic aperture radar. *Geophysical Research Letters*, 49(12):e2022GL098686, 2022.

[23] Ramon Torres, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, BjÖrn Rommen, Nicolas Floury, Mike Brown, et al. Gmes sentinel-1 mission. *Remote sensing of environment*, 120:9–24, 2012.

[24] Lars MH Ulander, M Lundberg, W Pierson, and A Gustavsson. Change detection for low-frequency sar ground surveillance. *IEE Proceedings-Radar, Sonar and Navigation*, 152(6): 413–420, 2005.

[25] Chen Wang, Alexis Mouche, Pierre Tandeo, Justin E Stopa, Nicolas Longépé, Guillaume Erhard, Ralph C Foster, Douglas Vandemark, and Bertrand Chapron. A labelled ocean sar imagery dataset of ten geophysical phenomena from sentinel-1 wave mode. *Geoscience Data Journal*, 6 (2):105–115, 2019.

[26] Chen Wang, Pierre Tandeo, Alexis Mouche, Justin E Stopa, Victor Gressani, Nicolas Longépé, Douglas Vandemark, Ralph C Foster, and Bertrand Chapron. Classification of the global sentinel-1 sar vignettes for ocean surface process studies. *Remote Sensing of Environment*, 234: 111457, 2019.

[27] Yi-Ran Wang and Xiao-Ming Li. Derivation of sea surface wind directions from terrasar-x data using the local gradient method. *remote sensing*, 8(1):53, 2016.

[28] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018.

[29] George S Young, TN Sikora, and NS Winstead. Use of synthetic aperture radar in finescale surface analysis of synoptic-scale fronts at sea. *Weather and forecasting*, 20(3):311–327, 2005.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Methods section.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Discussion section.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...

(a) Did you state the full set of assumptions of all theoretical results? [N/A]
(b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Instruction for reproduction are provided in Methods section, as well as a statement on data availability for the unlabeled data. Since the project is ongoing the code and labeled datasets have not been released yet. However, we fully intend to release both once the project is concluded.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Methods section.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] Due to time constraints, repeating the experiments has not been possible to derive error bars. Experiments are still ongoing and this will be added as soon as possible.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Methods section.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]