Naval Information Warfare Center PACIFIC

TECHNICAL DOCUMENT 3428
NOVEMBER 2023

Representations for Low-Budget Synthetic Aperture Radar (SAR) Active Learning

Martin T. Jaszewski
Jonathan K. Sato
Julian Y. Raheema
NIWC Pacific

Naval Information Warfare Center (NIWC) Pacific
San Diego, CA 92152-5001

This page is intentionally blank.

TECHNICAL DOCUMENT 3428
NOVEMBER 2023

Representations for Low-Budget Synthetic Aperture Radar (SAR) Active Learning

Martin T. Jaszewski
Jonathan K. Sato
Julian Y. Raheema

NIWC Pacific

Administrative Notes:
This document was authored in September 2023. It was approved through the Release of Scientific and Technical Information (RSTI) process in October 2023, and formally published in the Defense Technical Information Center (DTIC) in November 2023.

NIWC Pacific
San Diego, CA 92152-5001

NIWC Pacific
San Diego, California 92152-5001

P.M. McKenna, CAPT, USN
Commanding Officer

M.J. McMillan
Executive Director

ADMINISTRATIVE INFORMATION

The work described in this report was performed by the Naval Information Warfare Center Pacific (NIWC Pacific), San Diego, CA. The NIWC Pacific Naval Innovative Science and Engineering (NISE) Program provided funding for this applied research project.

ACKNOWLEDGMENTS

We would like to thank the NIWC Pacific High Performance Computing Center for providing cluster computing resources to support this work.

Editor: GG/MM

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

Computer vision models for synthetic aperture radar (SAR) imagery have the potential to enable round-the-clock, all-weather situational awareness for deployed Navy assets. To reach this potential they require sufficient labeled training data. However, labeling of maritime objects in SAR imagery is expensive due to the cost of specialized expertise to interpret imagery and label objects of interest.

Active learning (AL) aims to address this challenge by selecting a subset of the most informative data to label. Recent works for traditional computer vision problems such as image classification and object detection [1, 5, 7, 9] have reported success in meeting model performance targets with less labeled data. However, most of these methods rely on large initial sets of thousands of labeled images to work properly. Without the initial set, model performance tends to remain similar to random sampling of labeled images, which is unacceptable for practical applications.

Recently, self-supervised learning (SSL) has made significant advances, particularly in the area of contrastive learning [20-27]. SSL employs well-designed pretext tasks not directly related to the intended downstream task which induce the model to learn useful features without labeled data. Pretraining with self-supervision can be an effective means to prepare a model when the curated training set is small. However, SSL is not a substitute for supervised learning—some labeled data is still needed to train the model for the downstream task.

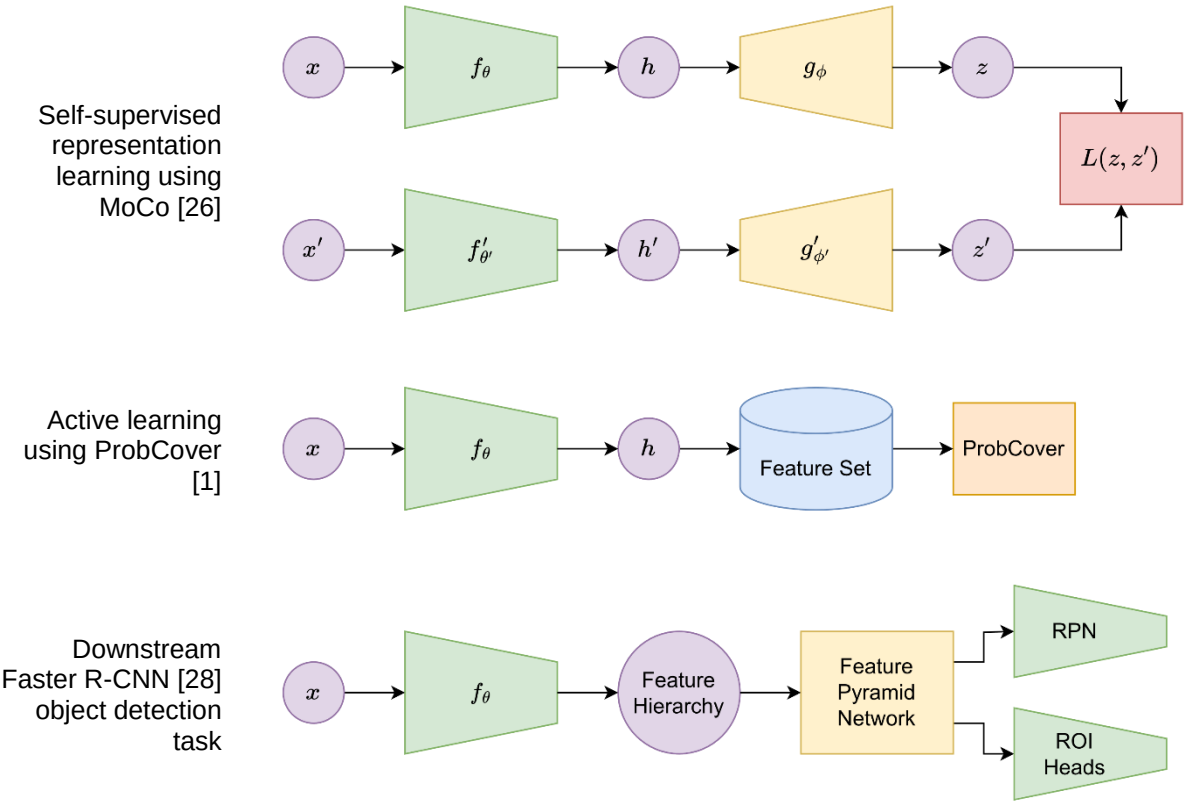

Figure 1. Representations learned through self-supervised learning can be transferred to active learning algorithms such as ProbCover [1] in addition to downstream tasks.

A subset of AL has recently emerged to directly address "low-budget" scenarios in which the number of labeled examples is very low. For our computer vision problems, we consider low-budget to be on the order of hundreds of labeled images. Several works [1-4] frame the low-budget AL task as a cluster sampling problem. In this context, the set of unlabeled images are represented as points in feature space. For models that have learned effective representations, visually similar images tend to cluster together in feature space, and dissimilar images are spread apart. The AL algorithm samples points from each cluster using strategies such as core-set [5], typicality [2], and max-cover [1]. The rationale for these approaches is that sampling from diverse clusters in feature space yields visually diverse images in input space. Low-budget AL algorithms rely upon robust features learned before data labeling, so they generally rely upon representations learned through SSL.

In this paper, we qualitatively assess the quality of feature representations for low-budget active learning applied to SAR images in the maritime domain. We consider features learned on images from three different source datasets with varying levels of domain mismatch to the target SAR maritime dataset. We visually compare the features and the images chosen by ProbCover [1], a state-of-the-art low-budget AL algorithm. This work provides preliminary qualitative results to motivate further quantitative analysis of low-budget AL for SAR computer vision.

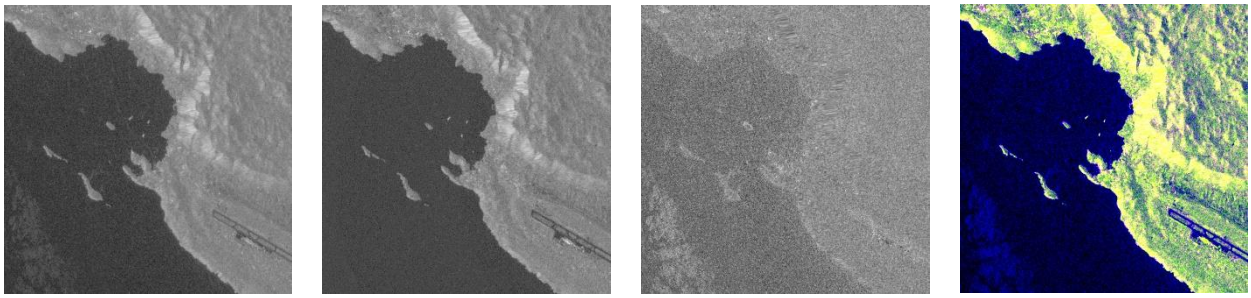

Figure 2. Color composite created from Sentinel-1 VV and VH polarimetric channels.

2. RELATED WORK

Existing works on low-budget active learning focus more on the cluster sampling algorithms and less on the quality of feature representations. In [3], the authors propose single- and multi-batch k-means algorithms to sample the images and provide an extensive ablation study, but they do not examine the effects of domain mismatch. The ProbCover algorithm [1] is compared to the Core-Set algorithm [5] across several datasets, but domain adaptation is not discussed. In [2], theoretical and empirical results agree on the effectiveness of typicality sampling in the low-budget regime, but there is no discussion on feature representation strength. The authors of [4] report effective domain adaptation between photos from three different domains of the Office-31 dataset [31], but they did not look at other image modalities.

Active learning algorithms for object detection tend to have large initial budgets. The authors of [8] compare six AL algorithms for object detection in which the initial budget ranges from 500 to 5,000 images. In [7] the initial set is 1,180 images.

Very few published works exist on low-budget AL for SAR computer vision. Most published works on AL for SAR investigate image classification [10] using the MSTAR dataset [19]. The authors in [13] and [11] report success for AL with budget sizes in the hundreds with a CNN-based variational autoencoder and SimCLR [27] respectively, but the MSTAR dataset is already small and not costly to label. The authors of [15] use a cosine distance-based diversity selection method and apply it to semantic segmentation of TerraSAR-X images using budget sizes from 200 to 1,000 images. In [12], the authors compare AL methods for object detection in xView3-SAR [17] images. Of these works, [12] and [15] are most closely related to this technical document, but no single work studies the role of representation quality for low-budget AL in SAR.

3. METHOD

3.1 PROBLEM FORMULATION

We consider the pool-based approach to active learning in which the AL algorithm chooses without replacement from an unlabeled target dataset X_U of fixed size. In this work we focus on the simplest case in which a single query expends the entire labeling budget B. This yields a labeled dataset $\{X_L, Y_L\}$, and the unlabeled dataset is reduced to $X_U = \{X_U - X_L\}$ where $|X_L| = B$ and $|\cdot|$ denotes cardinality.

We follow the general low-budget AL framework described in [1-4]. A model $f(\cdot)$ is trained on a source dataset through supervised, self-supervised, or other approaches. For each image $x_U \in X_U$ of width W, height H, and channels C we obtain the feature vector $h_U = f(x_U)$ of dimension D where $D \ll WHC$. This yields a set of D-dimensional feature vectors, one for each image in X_U. Each feature vector is a point in feature space. Under this problem setting, the task of the active learner is to select B points that form a representative subset of the unlabeled set in feature space.

3.2 DATASET PREPARATION

In this work we are interested in the xView3-SAR [17] maritime object detection dataset. xView3-SAR is the largest publicly available dataset for SAR ship detection research to date. The dataset consists of Sentinel-1 [35] image scenes with bounding box annotations for maritime objects.

We adapt the xView3-SAR dataset for active learning by filtering out unwanted areas of dataset image scenes and choosing appropriate input representations for extracting features from each image.

3.2.1 Sample Filtering

In most published works on AL, every sample in the unlabeled set is a potentially valid choice for labeling. For image classification datasets such as MSTAR [19] or ImageNet [18], this is true since each image is assigned at least one classification label. For each image in curated object detection datasets such as COCO [32] or Pascal VOC [33], each image is valid and contains one or more objects of interest. However, for large-scale satellite image datasets, objects of interest may be unevenly distributed or even absent in some areas.

For xView3-SAR, the vast majority of the image scene area contains no maritime objects of interest and/or invalid data. Another consideration for satellite image datasets like xView3-SAR is that each image scene is too large to be processed whole by currently available graphics processing units (GPUs). To handle these issues, we split the large image scenes into small chips and filtered out irrelevant chips.

We found that many chips contained no-data (invalid) pixel values, large areas of land, and empty expanses of water, so we designed filters to reject each of these categories of undesirable image content. By removing these unwanted chips from the unlabeled set prior to active learning, we improved computational efficiency by reducing the search space for the AL algorithm. We selected 155 image scenes from the xView3-SAR training set for training and all 50 validation scenes for validation. Table 1 summarizes the results of filtering image chips prior to active learning. Using a non-overlapping grid sampling pattern and a chip size of 256×256 pixels yielded 6,922,000 chips for training and 549,040 for validation. After applying the three-step filtering process, the size of the training set was reduced by 153×.

Table 1. Effects of filtering xView3-SAR dataset samples.

	Number of samples remaining after applying:			
Dataset split	Grid sampling	No-data filter	Land filter	Morph. filter
Training	6,922,000	1,288,321	968,977	45,063
Validation	549,040	351,414	224,522	18,800

3.2.1.1 No-Data and Land Filters

Areas of no-data pixel values pose a problem in the process of making full use of the dataset. When the source Sentinel-1 images were preprocessed for the xView3-SAR dataset, a reprojection step introduced no-data regions around the borders of the image. When creating a variation of the dataset by dividing up each image into smaller resolution images, areas of no-data can take up a significant portion of the new, smaller image. These areas are not informative to a detection algorithm and even reduce performance as areas of no-data will be expected when applied to new data. The no-data filter rejects a chip if the number of no-data pixels (indicated by an intensity value of -32768) exceeds 10% of the chip area.

Image chips of littorial regions contain valuable background context and negative examples such as rocks and islets, as well as human infrastructure such as docks, cranes, and other objects resembling ships. However, large expanses of land offer diminishing value to models for maritime use. We created landmasks using shoreline polygons from the OpenStreetMap project [34], an open-source project that maintains an up-to-date map of the world including land masses and bodies of water. We configured the land filter to reject a chip if the number of land pixels exceeds 50% of the chip area.

3.2.1.2 Morphological Filter

After reducing the number of images for consideration by the no-data and landmask filters, images with large areas of water remained. However, not every image is useful as most contain no objects of interest. To filter out these empty images, we employ a morphological-based filter.

Figure 3 illustrates the implementation of the morphological filter. First, we created a binary image by thresholding the VH polarimetric channel by a fixed value that will make sure all objects of interest are kept in the binary image. This intensity value was found by sampling multiple objects of interest as they all had similar reflectivity. After creating this binary image, we applied morphological opening to remove any objects too small to be of interest while retaining the size of the original thresholded objects. Next, we applied convolution with a 3×3 kernel of ones to identify objects that were large enough to be maritime objects or land features. If the resulting binary mask image did not contain any true-valued pixels, the chip is rejected.

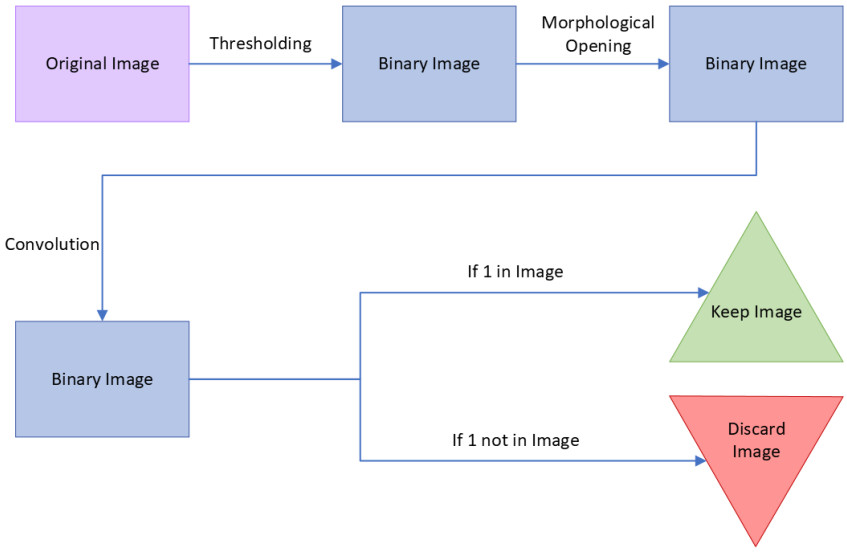

Figure 3. Flow chart of morphological filter.

3.2.2 Input Representation

Images from xView3-SAR are Sentinel-1 ground range detected (GRD) data products consisting of dual-polarization VV (co-polarized vertical transmit, vertical receive) and VH (cross-polarized vertical transmit, horizontal receive) polarimetric rasters. The preprocessing steps described in [17] yield backscatter pixel values in logarithmic scale (decibels).

In this work, we extract features from the xView3-SAR target dataset using models trained from three different source datasets. Two models were trained on RGB color images, and the third model was trained on dual-polarization Sentinel-1 images. Since xView3-SAR images are two-channel (dual-polarization), to extract features using models configured for three-channel color images we created RGB color composite images from the VV and VH channels as described in Table 2 (see Figure 2 for a visual example).

Table 2. xView3-SAR color composite details.

Color channel	Source raster	Mean	Standard deviation
Red	VV	-15 dB	10 dB
Green	VH	-23 dB	11 dB
Blue	VV-VH	-2 dB	15 dB

We used log-scale normalization parameters from [36].

3.3 REPRESENTATION LEARNING

We used ResNet-50 [29] models with three different sets of weights to extract feature vectors for each image chip. Features are taken from the output of the penultimate layer of the network immediately preceding the classification head. The three sets of weights for the ResNet-50 networks were learned from three different image datasets: ImageNet-1K [18], SSL4EO-S12 [16] Sentinel-2, and SSL4EO-S12 Sentinel-1. These source datasets represent varying levels of domain mismatch in

terms of sensor type, sensor perspective, and scene background environment. Table 3 summarizes the key differences between the source datasets and the xView3-SAR target dataset.

Table 3. Source and target image dataset differences.

Dataset	Sensor type	Sensor perspective	Scene environment
ImageNet-1K [18]	digital cameras	varies	varies
SSL4EO-S12 [16] Sentinel-2	multispectral EO	overhead	terrestrial
SSL4EO-S12 [16] Sentinel-1	SAR	overhead	terrestrial
xView3-SAR [17]	SAR	overhead	maritime

The ImageNet-1K model was trained using fully supervised learning on the ILSVRC 2012 (ImageNet-1K) dataset. ImageNet-1K is an image classification dataset containing 1,281,167 digital camera images of 1,000 classes of objects in everyday contexts. The model training procedure follows the protocol detailed in [37].

The remaining two models were trained using self-supervised learning on the SSL4EO-S12 dataset, a multimodal multitemporal satellite image dataset for unsupervised and self-supervised learning from remote sensing imagery. This dataset contains 2640×2640 meter patches from 251,079 global locations at four seasonally varying times from Sentinel-1 [35] SAR and Sentinel-2 [38] multispectral electro-optical (EO) satellite-based sensors. For this work, we were interested in self-supervised models pretrained using the MoCo [26] contrastive learning framework on RGB data from B4 (red), B3 (green), and B2 (blue) bands of Sentinel-2 patches and VV and VH polarimetric bands of Sentinel-1 patches.

Figure 4 shows example images from the ImageNet-1K and SSL4EO-S12 source datasets and the xView3-SAR target dataset.

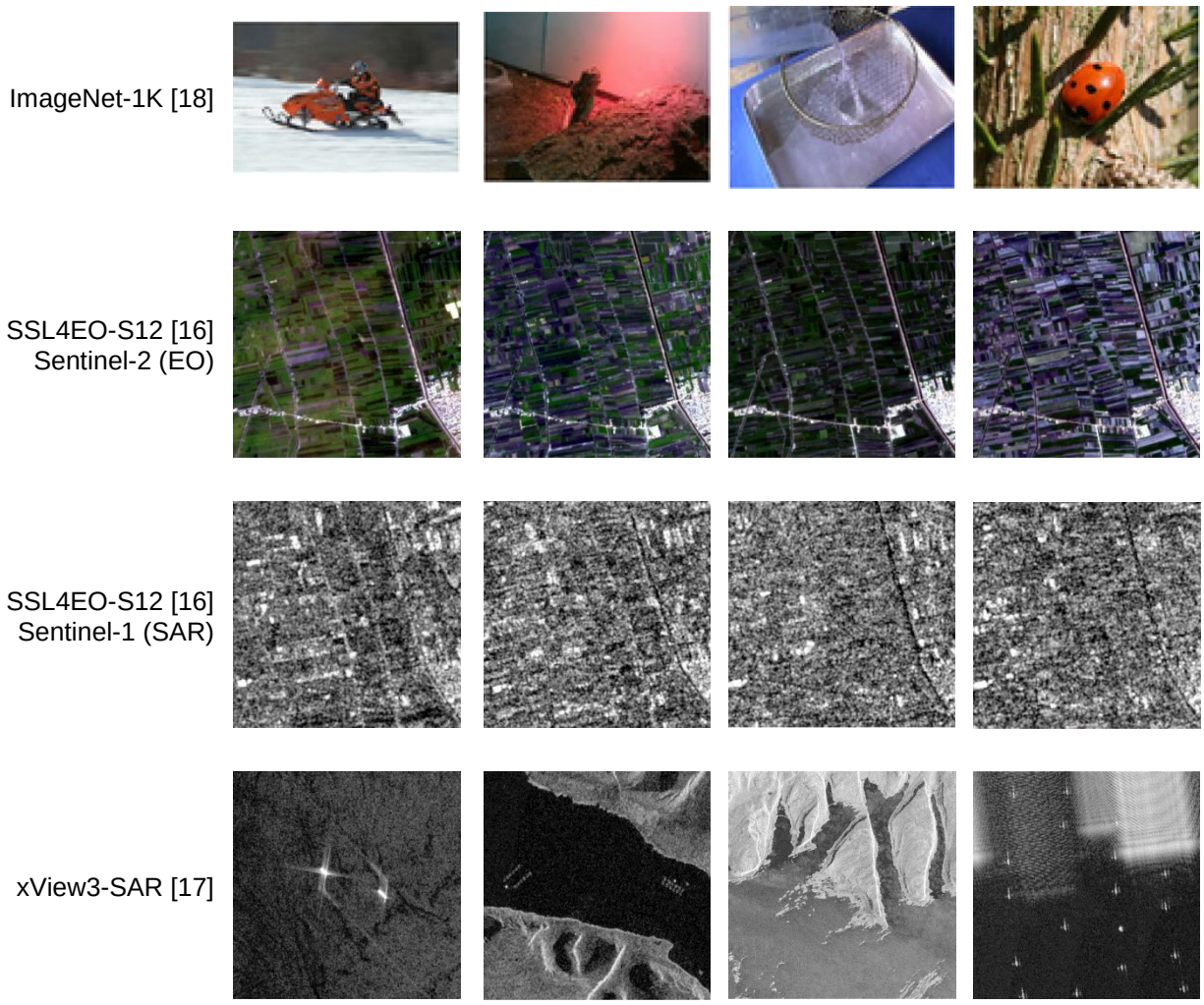

Figure 4. Example images from the ImageNet-1K and SSL4EO-S12 datasets used to train ResNet-50 models to extract features from xView3-SAR images.

3.4 ACTIVE LEARNING AS CLUSTER SAMPLING

ProbCover [1] is a sampling method that uses the geometry of the data representation for selecting the most meaningful samples. Given a set of feature vectors representing the unlabeled set X_U and a budget size B, ProbCover seeks to maximize coverage of the points in feature space, constrained by a fixed ball radius δ for each of B balls. Images corresponding to the central point of each ball are selected for labeling and training the task model.

4. EXPERIMENTS

4.1 EXPERIMENT CONFIGURATION

4.1.1 Feature Extraction

The ResNet-50 models for the ImageNet-1K and SSL4EO-S12 datasets were pretrained using the protocols described in [18] and [16] respectively.

We prepared the input representations of the xView3-SAR images according to the procedure described in section 3.2.2. We extracted features from the final bottleneck layer of the ResNet-50 models immediately before the classification head. For ResNet-50, this feature vector is 2,048-dimensional. On a single NVIDIA A100 GPU, it takes approximately 7 minutes to extract features from all 45,063 chips in the unlabeled set.

4.1.2 Active Learning

Due to scale differences in the extracted feature vectors, we varied the covering ball radius δ for the ProbCover algorithm across the three source datasets. We used δ = 5, 10, and 300 for ImageNet-V1, SSL4EO-S12 Sentinel-2, and SSL4EO-S12 Sentinel-1 datasets respectively. Selecting the top-16 image chips from the 45,063 chips in the unlabeled set took approximately 30 seconds.

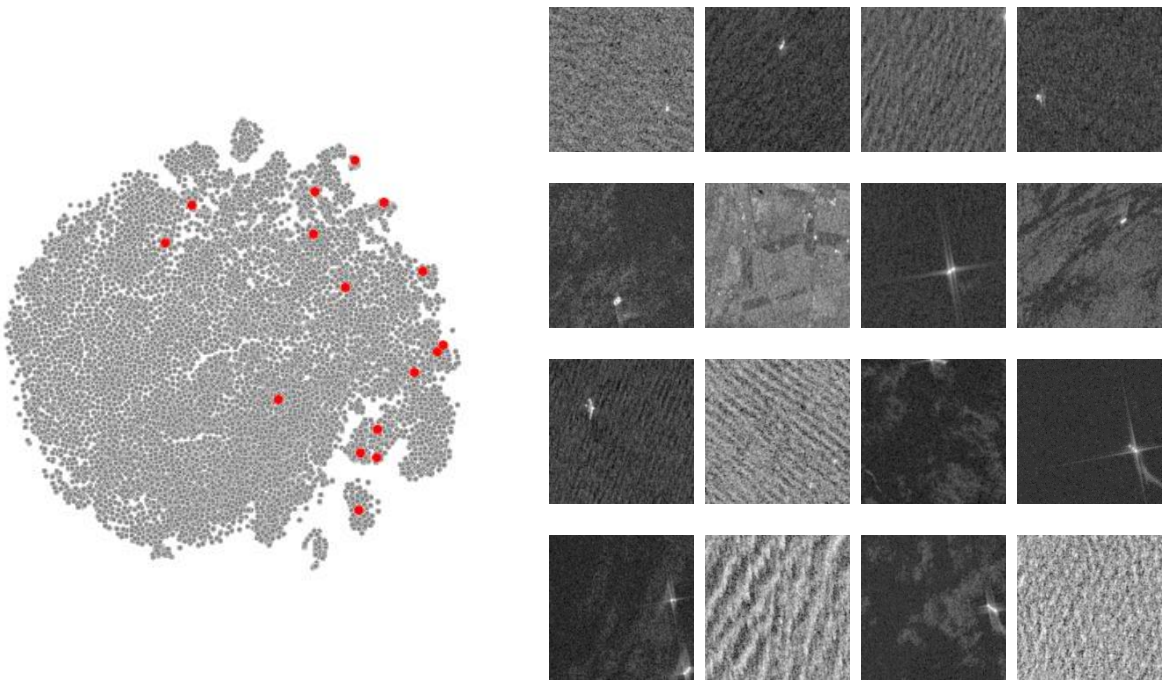

Figure 5. (left) t-SNE projected features extracted from xView3 training samples using the ImageNet-1K V2 model and (right) top-16 images selected.

4.2 RESULTS ON XVIEW3-SAR

Figures 5 through 7 (left side) illustrate the grouping of images in projected feature space. We used t-distributed stochastic neighbor embedding (t-SNE) [30] to project the 2,048-dimensional points to

two-dimensional space for visualization. Figures 5 through 7 (right side) show the top-16 image chips selected by the ProbCover algorithm given the features extracted by the corresponding model.

For ImageNet-1K extracted features (Figure 5), the data points are grouped into one large cluster and a few smaller clusters. The relative lack of visual diversity between the selected chips suggests that features learned from common digital camera photos may not effectively discriminate between xView3-SAR image chips to select representative examples.

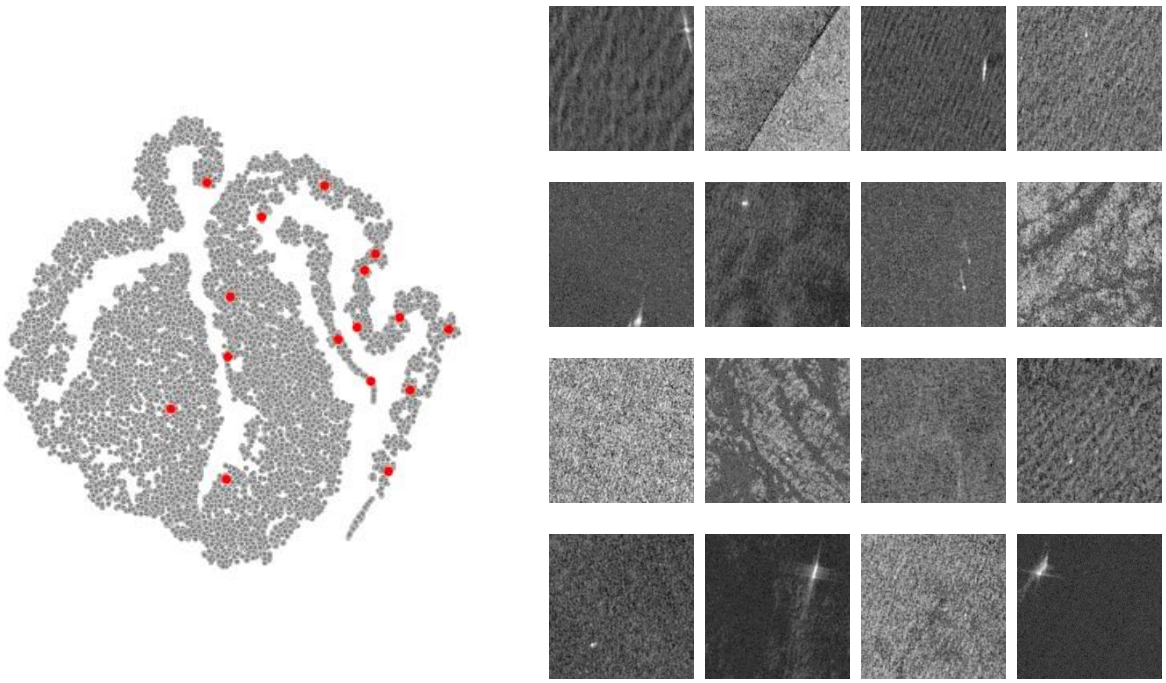

Figure 6. (left) t-SNE projected features extracted from xView3 training samples using the SSL4EO-S12 Sentinel-2 model and (right) top-16 images selected.

We observed significantly more distinct clusters for SSL4EO-S12 Sentinel-2 extracted features (Figure 6), suggesting that useful features transfer from electro-optical to SAR modalities. The two sensors have similar spatial resolution and look angles which encourage scale consistency of features. Selected chips capture a wide range of background appearance variations which we predict will encourage the task object detection model to learn foreground-background discrimination more quickly.

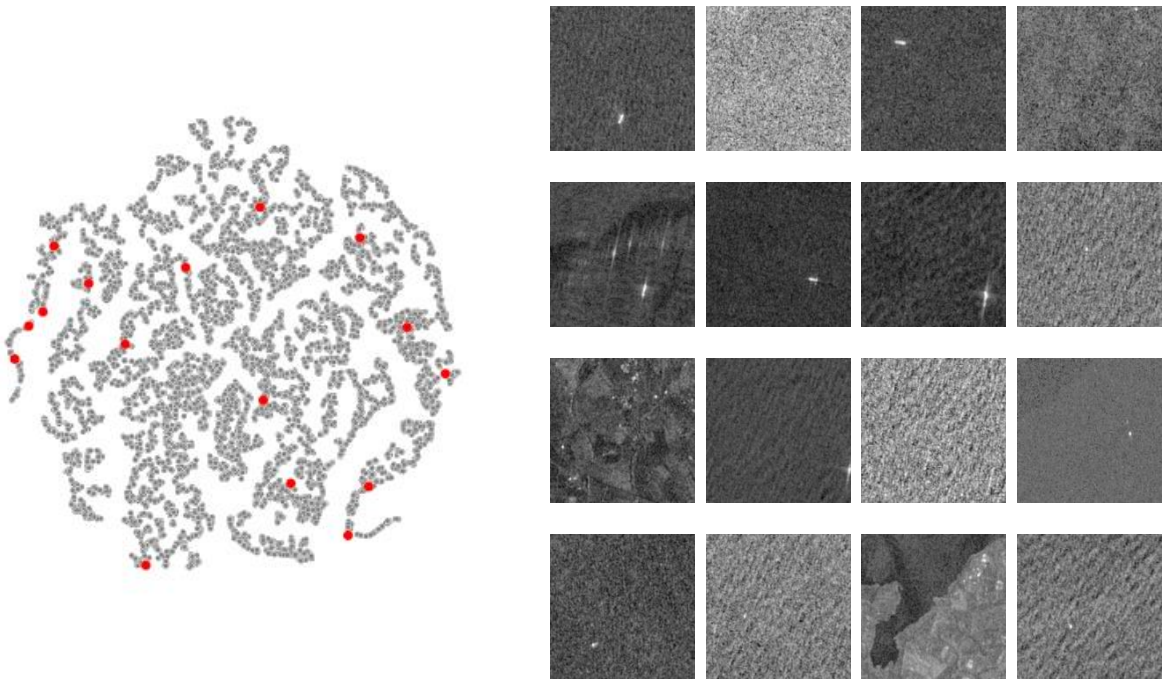

Figure 7. (left) t-SNE projected features extracted from xView3 training samples using the SSL4EO-S12 Sentinel-1 model and (right) top-16 images selected.

Features extracted from the SSL4EO-S12 Sentinel-1 model (Figure 7) are split into many small but distinct clusters, much more than the previous two models. This enables the AL algorithm to choose chips that represent a wide range of inter- and intra-class maritime object and background variations. However, the top-16 examples do not appear as visually distinct as those from the SSL4EO-S12 Sentinel-2 model. For future quantitative studies, we intend to look at a larger sample of selected chips and compare our visual qualitative assessments to performance metrics on the task model.

6. DISCUSSION AND CONCLUSION

In this paper, we compared the quality of feature representations for low-budget active learning for SAR maritime domain applications. We visualized features learned by models trained on images with varying levels of domain mismatch to the target SAR maritime dataset. We also visually compared images chosen from feature space using the ProbCover AL algorithm.

Our limited qualitative results support the hypothesis that using features learned from datasets from similar domains as the target dataset enable the AL algorithm to select more visually diverse images. We intend to supplement these preliminary qualitative results with further quantitative analysis to develop and impove low-budget AL algorithms for SAR computer vision.

6.1 LIMITATIONS

This work was a limited qualitative study of the effects of feature representations for low-budget SAR AL. The three models shared the same ResNet-50 architecture, but the training protocols differed substantially. The ImageNet-1K model was trained using fully supervised learning with labels and a standard cross-entropy loss, whereas the SSL4EO-S12 models were trained using the MoCo self-supervised learning framework. The small AL budget of only 16 images should be increased for more thorough visual comparison in the future. Nonetheless, the initial results presented in this technical document suggest future quantitative assessment and development of algorithms tailored to SAR datasets and models.

6.2 FUTURE WORK

We intend to develop pretext tasks designed with SAR and satellite imagery characteristics in mind so that we can apply SSL directly to SAR maritime imagery datasets and learn features relevant to this domain. We plan to test ProbCover and other cluster sampling based AL algorithms on these features so we can fully evaluate the utility of low-budget AL for SAR maritime computer vision applications.

This page is intentionally blank.

REFERENCES

1. Yehuda, Ofer; Avihu Dekel; Guy Hacohen; Daphna Weinshall [2022]: Active Learning Through a Covering Lens, arXiv: 2205.11320 [cs.LG]
2. Hacohen, Guy; Avihu Dekel; Daphna Weinshall [2022]: Active Learning on a Budget: Opposite Strategies Suit High and Low Budgets, arXiv: 2202.02794 [cs.LG]
3. Pourahmadi, Kossar; Parsa Nooralinejad; Hamed Pirsiavash [2022]: A Simple Baseline for Low-Budget Active Learning, arXiv: 2110.12033 [cs.CV]
4. Mahmood, Rafid; Sanja Fidler; Marc T. Law [2023]: Low Budget Active Learning via Wasserstein Distance: An Integer Programming Approach, arXiv: 2106.02968 [cs.LG]
5. Sener, Ozan & Silvio Savarese [2018]: Active Learning for Convolutional Neural Networks: A Core-Set Approach, arXiv: 1708.00489 [cs.ML]
6. Lyu, Mengyao; Jundong Zhou; Hui Chen; Yijie Huang; Dongdong Yu; Yaqian Li; Yandong Guo; Yuchen Guo, Liuyu Xiang; Guiguang Ding [2023]: Box-Level Active Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
7. Mi, Peng; Jianghang Lin; and Zhou, Yiyi and Shen, Yunhang and Luo, Gen and Sun, Xiaoshuai and Cao, Liujuan and Fu, Rongrong and Xu, Qiang and Ji, Rongrong [2022]: Active Teacher for Semi-Supervised Object Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
8. Yu, Weiping; Sijie Zhu; Taojiannan Yang; Chen Chen [2022]: Consistency-Based Active Learning for Object Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
9. Kim, Kwanyoung; Dongwon Park; Kwang In Kim; Se Young Chun [2021]: Task-Aware Variational Adversarial Active Learning, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
10. Andrews, Seth [2023]: Active learning for target detection and classification in SAR imagery, Algorithms for Synthetic Aperture Radar Imagery XXX PC12520.
11. Brown, Jason; Riley O'Neill; Jeff Calder; Andrea L. Bertozzi [2023]: Utilizing contrastive learning for graph-based active learning of SAR data, Algorithms for Synthetic Aperture Radar Imagery XXX 125200I.
12. Sato, Jonathan; Julian Raheema; Martin Jaszewski [2022]: Evaluating active learning methods for synthetic aperture radar maritime object detection, Artificial Intelligence and Machine Learning in Defense Applications IV 12276.
13. Miller, Kevin; John Mauro; Jason Setiadi; Xoaquin Baca; Zhan Shi; Jeff Calder; Andrea L. Bertozzi [2022]: Graph-based Active Learning for Semi-supervised Classification of SAR Data, arXiv: 2204.00005 [cs.LG]
14. Zhao, Siyuan; Ying Luo; Tao Zhang; Weiwei Guo; Zenghui Zhang [2022]: Active Learning SAR Image Classification Method Crossing Different Imaging Platforms, IEEE Geoscience and Remote Sensing Letters 19.
15. Wang, Tengchuan; Yuanxiang Li; Huilin Xiong [2014]: A novel locally active learning method for SAR image classification, 2014 IEEE Geoscience and Remote Sensing Symposium.

16. Wang, Yi; Nassim Ait Ali Braham; Zhitong Xiong; Chenying Liu; Conrad M Albrecht; Xiao Xiang Zhu [2023]: SSL4EO-S12: A Large-Scale Multi-Modal, Multi-Temporal Dataset for Self-Supervised Learning in Earth Observation, arXiv: 2211.07044 [cs.CV]

17. Paolo, Fernando; Tsu-ting Tim Lin; Ritwik Gupta; Bryce Goodman; Nirav Patel; Daniel Kuster; David Kroodsma; Jared Dunnmon [2020]: xView3-SAR: Detecting Dark Fishing Activity Using Synthetic Aperture Radar Imagery, Advances in Neural Information Processing Systems 35.

18. Russakovsky, Olga; Jia Deng; Hao Su; Jonathan Krause; Sanjeev Satheesh; Sean Ma; Zhiheng Huang; Andrej Karpathy; Aditya Khosla; Michael Bernstein; Alexander C. Berg; Li Fei-Fei [2015]: ImageNet Large Scale Visual Recognition Challenge, arXiv: 1409.0575 [cs.CV]

19. AFRL & DARPA; Moving and stationary target acquisition and recognition (MSTAR) dataset. https: //www.sdms.afrl.af.mil/index.php?collection=mstar. Accessed: 2021-07-10.

20. Balestriero, Randall; Mark Ibrahim; Vlad Sobal; Ari Morcos; Shashank Shekhar; Tom Goldstein; Florian Bordes; Adrien Bardes; Gregoire Mialon; Yuandong Tian; Avi Schwarzschild; Andrew Gordon Wilson; Jonas Geiping; Quentin Garrido; Pierre Fernandez; Amir Bar; Hamed Pirsiavash; Yann LeCun; Micah Goldblum [2023]: A Cookbook of Self-Supervised Learning, arXiv: 2304.12210 [cs.LG]

21. Bardes, Adrien; Jean Ponce; Yann LeCun [2022]: VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning, arXiv: 2105.04906 [cs.CV]

22. Zbontar, Jure; Li Jing; Ishan Misra; Yann LeCun; Stéphane Deny [2021]: Barlow Twins: Self-Supervised Learning via Redundancy Reduction, arXiv: 2103.03230 [cs.CV]

23. Koohpayegani, Soroush Abbasi; Ajinkya Tejankar; Hamed Pirsiavash [2020]: CompRess: Self-Supervised Learning by Compressing Representations, arXiv: 2010.14713 [cs.CV]

24. Grill, Jean-Bastien; Florian Strub; Florent Altché; Corentin Tallec; Pierre H. Richemond; Elena Buchatskaya; Carl Doersch; Bernardo Avila Pires; Zhaohan Daniel Guo; Mohammad Gheshlaghi Azar; Bilal Piot; Koray Kavukcuoglu; Rémi Munos; Michal Valko [2020]: Bootstrap your own latent: A new approach to self-supervised Learning, arXiv: 2006.07733 [cs.LG]

25. Chen, Xinlei; Haoqi Fan; Ross Girshick; Kaiming He [2020]: Improved Baselines with Momentum Contrastive Learning, arXiv: 2003.04297 [cs.CV]

26. He, Kaiming; Haoqi Fan; Yuxin Wu; Saining Xie; Ross Girshick [2020]: Momentum Contrast for Unsupervised Visual Representation Learning, arXiv: 1911.05722 [cs.CV]

27. Chen, Ting; Simon Kornblith; Mohammad Norouzi; Geoffrey Hinton [2020]: A Simple Framework for Contrastive Learning of Visual Representations, arXiv: 2002.05709 [cs.LG]

28. Ren, Shaoqing; Kaiming He; Ross Girshick; Jian Sun [2017]: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6).

29. He, Kaiming; Xiangyu Zhang; Shaoqing Ren; Jian Sun [2020]: Deep Residual Learning for Image Recognition, arXiv: 2010.14713 [cs.CV]

30. van der Maaten, Laurens & Geoffrey Hinton [2008]: Visualizing Data using t-SNE, Journal of Machine Learning Research 9(86).

31. Saenko, Kate; Brian Kulis; Mario Fritz; Trevor Darrell [2010]: Adapting Visual Category Models to New Domains, European Conference on Computer Vision 11.

32. Lin, Tsung-Yi; Michael Maire; Serge Belongie; Lubomir Bourdev; Ross Girshick; James Hays; Pietro Perona; Deva Ramanan; C. Lawrence Zitnick; Piotr Dollár [2015]: Microsoft COCO: Common Objects in Context, arXiv: 1405.0312 [cs.CV].

33. Everingham, M.; S. M. A. Eslami; L. Van Gool; C. K. I. Williams; J. Winn; A. Zisserman, [2010]: The Pascal Visual Object Classes Challenge: A Retrospective, International Journal of Computer Vision 111(1).

34. OpenStreetMap contributors. (2017). Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org.

35. European Space Agency. (2023). Sentinel-1 Overview. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview.

36. TCR. (2021). Creating composite RGB images from Sentinel 1 channels. https://gis.stackexchange.com/questions/400726/creating-composite-rgb-images-from-sentinel-1-channels.

37. Vyrniotis, Vasilis. (2021). How to Train State-Of-The-Art Models Using TorchVision's Latest Primitives. https://pytorch.org/blog/how-to-train-state-of-the-art-models-using-torchvision-latest-primitives.

38. European Space Agency. (2023). Sentinel-2 Overview. https://sentinel.esa.int/web/sentinel/missions/sentinel-2/overview.

INITIAL DISTRIBUTION

84310	Technical Library/Archives	(1)
56220	M. Jaszewski	(1)
53624	J. Sato	(1)
56481	J. Raheema	(1)

Defense Technical Information Center
Fort Belvoir, VA 22060–6218 (1)

Naval Innovative Science and Engineering (1)

This page is intentionally blank.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-01-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
November 2023	Final	

4. TITLE AND SUBTITLE

Representations for Low-Budget SAR Active Learning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS

Martin T. Jaszewski
Jonathan K. Sato
Julian Y. Raheema
NIWC Pacific

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NIWC Pacific
53560 Hull Street
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

TD-3428

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Innovative Science and Engineering
53560 Hull Street
San Diego, CA 92152–5001

10. SPONSOR/MONITOR'S ACRONYM(S)

NISE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Active learning (AL) seeks to minimize annotation costs by selecting the most useful training data for deep machine learning models. However, until recently, the most successful active learning algorithms have required a large initial labeled set to work properly. This makes them impractical for labor-intensive labeling tasks such as object detection and modalities such as synthetic aperture radar (SAR). Improvements in extracting useful intrinsic features from unlabeled data through self-supervised learning have reduced the need for large amounts of labeled data. As a result, AL can focus on selecting a smaller subset of the best data to label for supervised learning on downstream tasks. Since annotation costs tend to be higher for these downstream tasks, it is crucial that the AL algorithm perform well under low annotation budgets. In this work, we study feature representations for AL in the low-budget regime. Our AL approach builds upon existing self-supervised representation learning methods by leveraging features learned from pretrained models and applying them to the SAR modality. Using these self-supervised learned features, we employ a diversity-based sampling strategy to select the examples for labeling under a low annotation budget. Qualitative results indicate that self-supervision with datasets well-matched to the target domain, when combined with cluster sampling-based AL algorithms, can lead to the selection of diverse samples.

15. SUBJECT TERMS

active learning, computer vision, machine learning, self-supervised learning, synthetic aperture radar, satellite imagery, object detection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Martin T. Jaszewski
U	U	U	SAR	26	19b. TELEPHONE NUMBER *(Include area code)* (619) 553-4256

This page is intentionally blank.

Approved for public release. Distribution is unlimited.

Naval Information Warfare Center (NIWC) Pacific
San Diego, CA 92152-5001