A STUDY ON THE EFFECT OF COMMONLY USED DATA AUGMENTATION TECHNIQUES ON SONAR IMAGE ARTIFACT DETECTION USING DEEP NEURAL NETWORKS

M. Orescanin, B. Harrington, D. Olson

Naval Postgraduate School
Monterey, CA, USA

M. Geilhufe, R. E. Hansen, N. Warakagoda

Norwegian Defense Research Establishment
Kjeller, Norway

ABSTRACT

This paper presents an empirical study that evaluates the impact of different types of augmentations on the performance of Deep Learning (DL) models for detecting imaging artifacts in Synthetic Aperture Sonar (SAS) imagery. Despite the popularity of using DL in the SAS community, the impact of augmentations that violate the geometry and physics of SAS has not been fully explored. To address this gap, we developed a unique dataset for detecting imaging artifacts in SAS imagery with DL and trained a Bayesian neural network with a ResNet architecture using widely used augmentations in DL for computer vision, as well as common augmentations used in the SAS literature. The study shows that augmentations that violate the geometry and imaging physics of SAS can negatively impact supervised classification, but can sometimes improve performance. Overall, the study provides important insights into the impact of different types of augmentations on the performance of DL models in SAS applications.

Index Terms— Synthetic Aperture Sonar, Deep Learning, Imaging Artifacts, Bayesian Deep Learning

1. INTRODUCTION

Deep learning (DL) has recently been applied to active sonar tasks such as target detection and seafloor classification [1, 2], but is challenging to implement in synthetic aperture sonar (SAS) due to sparse training data. Hence, a common technique is to use data augmentation to develop a viable model. Data augmentation techniques are domain specific and in the case of SAS, they could be synthetic through simulation with realistic imaging [3], augmentations of imagery through pseudo-coloring [2], or standard augmentations used in computer vision (e.g. translation, rotation, reflection). Williams [4] stated that image mirroring across-track or rotating images would violate physics of imaging, and mirroring along-track preserved the imaging physics. This was formalized in [1] with the suggestion that smaller range and along-track translations, within 0.5 meters relative to the image size of 5 meters, are acceptable. Others in the field followed these recommendations [2, 5]. Other augmentations were not explored, because it is unknown if they would negatively impact the performance of DL models. We address this gap by providing empirical evidence that augmentations violating unique geometry and imaging physics of SAS negatively impact supervised classification tasks. For that purpose we have developed a unique dataset for detecting imaging artifacts in SAS imagery with DL. Further, we consider widely used augmentations in DL for computer vision (e.g. rotation, zooming, contrast, cross-track mirroring) as well as common augmentations used in the SAS literature such as mirroring along-track. We train a Bayesian neural network using MC Dropout with a ResNet architecture for this task [6].

2. METHODOLOGY

2.1. Data used in the Study

Data were collected using the HISAS interferometric SAS carried by the HUGIN-HUS AUV of the Norwegian Defence Research Establishment (FFI) [7] from several different locations with a variety of seafloor types. Using backprojection imaging and taking into account navigation correction, SAS images were formed. A separate physically perturbed dataset was created by introducing imaging artifacts (which degraded the focus, grating lobe level, and SNR of the images). Large SAS scenes were tiled into 300×300 pixel patches and our overall approach produces 84k training samples, 5k validation and testing samples each with class-balanced validation and testing splits. This is a supervised learning approach with multi-class classification task. Ablation studies of probabilistic models have been shown for this dataset in [8].

2.2. Models

In this project, we used a ResNet 20 architecture [9] with 20 2D convolutional layers, each with 16 filters, a kernel size of 5×5, and a stride of 1 with zero padding. We used MC Dropout, applied during both training and inference, by placing dropout layers after each activation layer within the residual block, as detailed by Nado et al. [10].

DOI til publisert versjon/DOI to published version: 10.1109/IGARSS52108.2023.10282626

Table 1. Data Augmentation Techniques

Data Augmentation	Physical	Magnitude
Flip Along Track	Yes	50%
Flip Across Track	No	50%
Rotate Small	Yes	$\pm$ 5°
Rotate Large	No	$\pm$ 30°
Zoom	No	$\pm$ 1.2 m
Translate	No	$\pm$ 1.2 m
Contrast	Yes	$\pm$ 20%

2.3. Computer Vision Data Augmentations

The key model comparisons studied here are "physical" versus "non-physical" data augmentations. Table 1 summarizes the various data augmentation techniques employed, which were studied separately.

The mirroring functions were randomly initialized by TensorFlow (TF) using the RandomFlip function. Half the time, the preprocessing layer would mirror or flip the image horizontal (along track) or vertical (across track) as appropriate. Because the image is created by collecting sonar inputs along the track, flipping an image along the track is physical. In contrast, flipping across track will cause a shadow and its originating feature to flip, which violates the physical geometry of the scene. Therefore, mirroring along the range is classified as non-physical.

Rotation was based on the beamwidth of 20° of the elements in the SAS system. This work used rotations smaller and larger than this beamwidth, namely 5° and 30°. If the rotation is significantly less than the beamwidth, the resulting augmentation is physical, and vice versa. Zoom was initialized with a maximum $\pm 20\%$, corresponding to 1.2 meters of zoom or 60 pixels. Zoom is considered a non-physical perturbation due to its effect on the speckles in a SAS image. Translate shifts is considered a physical perturbation, but in this implementation the missing space is filled with zeros, which is non-physical.

The contrast augmentation changes the relative variation of the signal around the mean intensity by up to 20% in either direction. This change in relative intensity could be comparable to viewing the same image from a different altitude or grazing angle [11]. A model trained with contrast could become more range-invariant to future test data. As this augmentation is reproducible, it is we consider it to be a physical augmentation.

2.4. Pseudo-Coloring Techniques

Additionally, several different pseudo-coloring techniques were applied. Pseudo-coloring is treated as a physical augmentation technique. All colormaps are from Matplotlib listed colormaps [12] with the exception of the colormap designed by Isaac Gerg [13]. The Gerg colormap was designed to maximize perceptual luminance for SAS images. The images were created by mapping the image intensity to the RGB domain specified by the colormap, interpolating if necessary. The Matplotlib colormaps were discretized with 256 values, whereas the Gerg colormap was interpolated to double precision floating point. The images were truncated at 60 dB of dynamic range before conversion to the colormap. All colormaps tested against are illustrated in Fig. 1 on an example image.

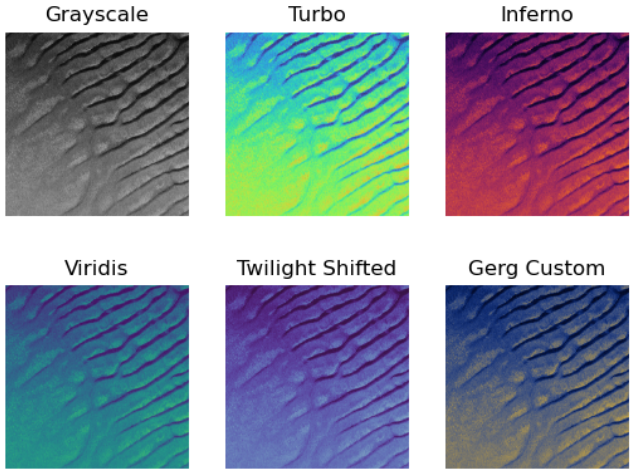

Fig. 1. Illustration of the colormaps considered.

2.5. Overall Model Development and Evaluation

There have been limited studies performed on image quality assessment for synthetic aperture images, so several different approaches were taken to determine which type of model works best for the problem. The base model was a ResNet 20v1 model [9], and that model was run using the MC dropout method [10] under fixed conditions over all augmentations studies in order to have a fair evaluation of results. Default model training included L2 regularization parameter of 0.001, dropout probability of 0.2, batch size of 128m, Adam optimizer with an initial learning rate of 0.001. We monitored for overfitting via early stopping strategy.

We focused on different forms of data augmentation, both computer vision techniques (see Table1) and pseudo-coloring (see Table 4). We applied five different colormaps to the MC Dropout model: Inferno, Twilight Shifted, Turbo, Viridis, and Gerg.

Analysis is conducted both in terms of standard metrics on the classification task (i.e., accuracy, precision, recall, f1-score) as well as the mean negative log likelihood metric (MNLL) over the test set [14]. This metric is defined as $\text{MNLL} = -\left(\frac{1}{N}\sum \log(p(y_c))\right)$, where N quantifies the sample size of the test dataset, and $p(y_c)$ is the probability the model assigns to the correct class label.

Dette er en postprint-versjon/This is a postprint version.
DOI til publisert versjon/DOI to published version: 10.1109/IGARSS52108.2023.10282626

3. RESULTS

We first demonstrate ResNet20 model baseline performance with no augmentations in training with accuracy of 0.867 and an MNLL of 0.563 on a balanced test set (see Table 2), against which we compare performance of augmentations. Further, we show results that demonstrate the effects of commonly used augmentation of along-track flip, versus across-track flip. Both accuracy and MNLL for flip along track improve over no augmentation performance while flip across track degrades relative to the no augmentation case, see Table 2.

Data Augmentation	Accuracy	MNLL
No Augmentation	86.7%	0.563
Flip Along Track	88.2%	0.385
Flip Across Track	81.0%	0.850
Zoom ($\pm 20\%$)	92.0%	0.297
Contrast ($\pm 20\%$)	88.6%	0.378
Rotate small ($\pm 5°$)	86.0%	0.474
Rotate large ($\pm 30°$)	86.7%	0.390
Translate (± 1.2 m)	88.9%	0.363

Table 2. Data Augmentation effect on the test accuracy and MNLL of the MC Dropout model. In the table we provide augmentations and their range as applied to the dataset.

We bring out an interesting impact of rotation since Table 2 suggests that large rotation enhances model performance which could be misleading. To show this we inspect the per-class performance of no augmentation and rotate large evaluations as given in Table 3.

	No Augmentation			Rotate large ($\pm 30°$)		
	prec	rec	f1	prec	rec	f1
No Artifact	0.77	0.86	0.81	0.69	0.95	0.80
Sound Speed	0.97	0.85	0.91	0.99	0.93	0.96
SNR	1.00	0.99	1.00	1.00	1.00	1.00
Yaw	0.75	0.76	0.76	0.86	0.59	0.70

Table 3. No Augmentation compared to a rotation of 30°. The overall accuracy is 0.867 for both augmentations.

By looking into precision and recall as metrics we can see that the per-class performance between no artifact and yaw is inversely impacted. Recall is reduced for yaw while precision is increased, which means that model does not detect many yaw artifacts but when it does it is correct. On the other hand, recall is very high with reduction in precision for the no artifact class. The model mis-classifies many images as no artifact. This could be explained by the fact that grating lobes always show up in the along track direction from bright objects, and rotation may violate this rule. If an object was big enough, some distortion would be observable in the along track direction, but not for small objects.

Zoom is a meaningful perturbation for large objects, because blurring and grating lobes can impact a variety of scatterer sizes, and the zoom augmentation can broaden the distribution of scatterer sizes present in the image. This may explain the increased performance of zoom, even though it alters the size of speckle blobs in the image. Zooming out also has the potential to act as a kind of despeckling filter, and may make the model less sensitive to imaging speckle. As discussed earlier, contrast is a physically meaningful perturbation since different grazing angles and rms terrain slope will have different contrast, and this augmentation increases the performance [11].

For the pseudo-color images, the working hypothesis was that generic Matplotlib colormaps would underperform grayscale, but the Gerg colormap would perform better. The process of applying colormaps to the images resulted in a loss of precision, reducing the amount of information available for analysis. This loss of precision could be crucial in distinguishing seafloor features and artifacts, particularly in regions of low signal intensity. Pseudo coloring also increased the model input features, which can lead to overfit with small datasets such as this one.

The overall results supported the hypothesis as all pseudo-coloring techniques performed worse than the grayscale original (see Table 4) . The Gerg colormap, which retained the original 32-bit precision, performed the best among the colormaps, but less than grayscale. Once again, we break out results for the yaw class of error, in Table 5. In our analysis we found that classification reports showed that errors in the "yaw" category significantly influenced the overall performance. Both the Gerg colormap and grayscale had higher f1 scores for yaw compared to other colormaps, 0.734 and 0.757 respectively. This suggests that the difference in precision between the original data and the colormaps played a substantial role in yaw classification and overall model performance. The results also indicated a correlation between no artifact and yaw classifications, suggesting that these categories were often confused by the model.

Colormap	Accuracy	MNLL
Grayscale	86.7%	0.563
Inferno	79.4%	0.800
Turbo	77.7%	0.969
Twilight Shifted	73.7%	1.381
Viridis	81.4%	0.856
Gerg	84.5%	0.715

Table 4. Effect of colormap on accuracy and MNLL

Dette er en postprint-versjon/This is a postprint version.
DOI til publisert versjon/DOI to published version: 10.1109/IGARSS52108.2023.10282626

Colormap	yaw (f1)	No Artifact (f1)
Grayscale	0.757	0.814
Inferno	0.609	0.713
Turbo	0.653	0.749
Twilight Shifted	0.669	0.669
Viridis	0.650	0.720
Gerg	0.734	0.785

Table 5. Yaw versus no artifact performance per colormap.

4. DISCUSSION AND CONCLUSION

We explored the effects of different data augmentation techniques on the performance of a Bayesian ResNet20 model in recognizing imaging artifacts in SAS imagery that were introduced via beamforming. The baseline performance of the model, without any augmentations, achieved an accuracy of 86.7% and an MNLL of 0.563 on this task. Overall, the computer vision perturbations that mesh well with the physics of imaging improve performance, and the perturbations that violate imaging physics degrade performance, except for the case of large rotation and zoom. The pseudo-coloring study aimed to evaluate the performance of different pseudo-color images compared to grayscale. The study found that all pseudo-coloring techniques performed worse than grayscale. We believe that the loss of precision during the pseudo-coloring process, combined with a much larger number of input features was responsible for the reduced performance. The Gerg colormap and grayscale showed better performance in classifying yaw errors, indicating the importance of precision in data representation. Additionally, a correlation was observed between no artifact and yaw classifications, suggesting confusion between these categories. By simulating diverse imaging conditions through perturbations in beamforming, we provide valuable insights for optimizing data augmentation in training of neural networks.

5. ACKNOWLEDGEMENTS

This research was supported by a grant from the Office of Naval Research (N-00014-22-WX-01861). The authors thank FFI's HUGIN-HUS operator group for collecting the data used in this study.

6. REFERENCES

[1] D. P. Williams, "On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery," *IEEE J. Ocean. Eng.*, vol. 46, no. 1, pp. 236–260, 2020.

[2] N. D. Warakagoda and Ø. Midtgaard, "Transfer-learning with deep neural networks for mine recognition in sonar images," in *Int. Conf. Underwat. Ac.*, 2020.

[3] D. Stewart, A. Kreulach, S. F. Johnson, and A. Zare, "Image-to-height domain translation for synthetic aperture sonar," *IEEE Trans. Geosci. Rem. Sens.*, vol. 61, pp. 1–13, 2023.

[4] D. P. Williams, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," in *23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2497–2502.

[5] T. S. Brandes, B. Ballard, S. Ramakrishnan, E. Lockhart, B. Marchand, and P. Rabenold, "Environmentally adaptive automated recognition of underwater mines with synthetic aperture sonar imagery," *J. Acoust. Soc. Am.*, vol. 150, no. 2, pp. 851–863, 2021.

[6] P. Ortiz, M. Orescanin, V. Petković, S. W. Powell, and B. Marsh, "Decomposing satellite-based classification uncertainties in large earth science datasets," *IEEE Trans. Geosci. Rem. Sens.*, vol. 60, pp. 1–11, 2022.

[7] R. E. Hansen, H. J. Callow, T. O. Sæbø, and S. A. V. Synnes, "Challenges in Seafloor Imaging and Mapping with Synthetic Aperture Sonar," *IEEE Trans Geosci. Rem. Sens.*, vol. 49, no. 10, pp. 3677–3687, 2011.

[8] M. Orescanin, B. Harrington, D. Olson, M. Geilhufe, R. E. Hansen, and N. Warakadogda, "Uncertainty Quantification with Deep Learning through Variational Inference with applications to Synthetic Aperture Sonar," in *Proc. Underwat. Acoust. Conf. 2023*, Kalamata, Greece, June 2023.

[9] K. He, X. Zhang, Sh. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE (CVPR)*, June 2016.

[10] Z. Nado et al., "Uncertainty Baselines: Benchmarks for uncertainty & robustness in deep learning," unpublished, 2021 [Online].

[11] A. P. Lyons, D. R. Olson, and R. E. Hansen, "Modeling the effect of random roughness on synthetic aperture sonar image statistics," *J. Acoust. Soc. Am.*, vol. 152, no. 3, pp. 1363–1374, sep 2022.

[12] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[13] I. Gerg, ," Private Communication, 2023.

[14] K. P. Murphy, *Probabilistic machine learning: an introduction*, MIT press, 2022.