Selection of Pre-training Datasets for Sonar Image Classification

Yannik Steiniger[1], Jose Luis Quinones Gonzalez[1], Dieter Kraus[2], Benjamin Lehmann[2]

Deep learning based computer vision models like convolutional neural networks (CNN) and Vision Transformer (ViT) are more and more applied for the automatic analysis of sonar images. Since sonar image datasets typically have a limited number of samples, transfer-learning is used to train these models. However, commonly used pre-training datasets, like ImageNet, have a large domain gap to sonar images, i.e., images in these two datasets are fundamentally different. The selection of the pre-training dataset and the related domain gap have shown to have an impact on the final performance of the model. In this work, different datasets are analysed for applying transfer-learning to deep learning models for the classification of sidescan sonar images. In addition, the study is conducted for shallow CNNs, deeper CNNs as well as ViT. We quantify the domain gap using a variational autoencoder (VAE) and the t-distributed stochastic neighbor embedding t-SNE and link these values to the classification performance of the models after fine-tuning. Our results show that while no dataset leads to an improvement of all models, the Fetal dataset works well for most investigated models, while ImageNet and its grayscaled version led to a worse performance.

KEY WORDS
~ Deep learning
~ Sidescan sonar
~ Transfer-learning
~ Computer vision
~ Sonar image classification
~ Pre-training

[1] German Aerospace Center, Institute for the Protection of Maritime Infrastructures, Bremerhaven, Germany
[2] HSB Hochschule Bremen – City University of Applied Sciences, Bremen, Germany
e-mail: yannik.steiniger@dlr.de

doi: 10.7225/toms.v14.n03.w03

Received: 26 May 2025 / Revised: 30 Jun 2025 / Accepted: 30 Jun 2025 / Published: 25 Jul 2025

1. INTRODUCTION

Surveying underwater areas is of great interest for the security of maritime infrastructures. However, due to the physical properties of water and turbidity of most sea areas, electromagnetic waves experience a high attenuation (Urick, 2013). Thus, instead of optical cameras, sonar systems, which transmit and receive acoustic waves, are used in the underwater domain. Sonar image data is typically captured using autonomous underwater vehicles (AUVs) or ships equipped with sidescan or synthetic aperture sonars. These vessels survey the seafloor following a pre-defined path to search for sunken objects of interest. Special signal processing routines transform the recorded acoustic signals to form intensity images of the seafloor which can be analysed by human operators or computer vision models, e.g., neural networks. However, collecting data is cumbersome and costly as it requires specialized personnel and equipment. Additionally, since the location of objects is in general unknown prior to scanning the seafloor and the ratio of object vs. seafloor in the captured data is low, datasets of relevant objects are very small. Although significant progress has been made in the field of sonar image classification, especially by developing specific deep learning models (Phung et al., 2019; Williams, 2021; Steiniger et al., 2022; Warakagoda and Midtgaard, 2024), the outcome of deep learning models remains limited due to the need for more training data.

In applications in which the available amount of data is limited, transfer-learning can be applied to improve the overall training of the models (Mensink et al., 2022). With this concept, a model is first pre-trained on a large source dataset, which does not necessarily represent the final task. In a second step the model is fine-tuned on the real target dataset. For classification, one of the most common pre-training dataset is ImageNet (Deng et al., 2009), containing over one million optical RGB images and 1,000 object classes. However, as shown in Figure 1, compared to sidescan sonar images, the samples from ImageNet are fundamentally different in terms of the imaging sensor, image content and resolution, resulting in a large domain gap between ImageNet and a sonar image dataset. Features that are learned by a deep learning model using ImageNet might be useless for the later classification of sonar images, e.g., features based on color. Researchers have shown that selecting pre-training datasets with a small domain gap to the target dataset can improve the performance of the model (Mensink et al., 2022). For sonar image classification, most models which are trained using transfer-learning use ImageNet as the source dataset (Warakagoda and Midtgaard, 2024; Sheffield et al., 2024), accepting a large domain gap and consequently potential suboptimal performance. To the best of our knowledge, there is no research about which dataset should be used for pre-training in the context of sonar image classification.

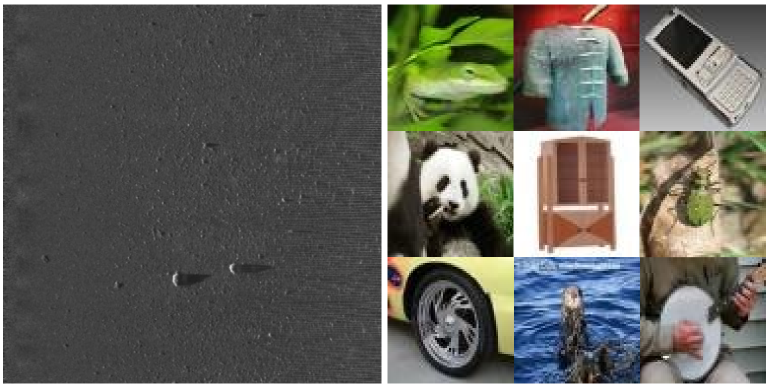


Figure 1. Sidescan sonar image spanning an area of 50 m×50 m with a resolution of 10 cm per pixel (left) and nine random images from the ImageNet dataset (right).

This paper presents a study on classification datasets, which, compared to ImageNet, are expected to have a smaller domain gap to a sidescan sonar image dataset. For measuring the domain gap we calculate the Euclidean distance in the latent space of a variational autoencoder (VAE) as well as the Euclidean distance in the t-distributed stochastic neighbor embedding (t-SNE) representation. We train different deep leaning models, ranging from a shallow convolutional neural network (CNN) used in our previous work (Steiniger et al., 2023) to deeper ones like ResNet-101 as well as Vision Transformer (ViT). The classification performance after fine-tuning on our own sonar image dataset is compared to the one of the networks trained from scratch. Our main finding is that none of the investigated datasets improves the performance of all deep learning models. Additionally, the domain gap is not the only parameter which influences the transfer-learning performance but also the size of the pre-training dataset.

The remaining of the paper is organized as follows: Section 2 introduces the datasets which are investigated in this work. Afterwards, Section 3 covers the measurement of the domain gap. The designed networks and training parameters are briefly explained in Section 4. In Section 5 the results of our experiments are presented. Finally, Section 6 closes the paper with a summary of the main findings and outlook to future work.

2. SELECTED DATASETS

Sidescan sonar data was collected over multiple sea trials in the time span from 2019 to 2023 using an Edgetech 2205 sidescan sonar mounted on a SeaCat AUV. The sonar image dataset build from these trials was already described in (Steiniger et al., 2023). It contains objects from the four classes *Tire*, *Rock*, *Cylinder* and *Wreck* as well as an additional *Background* class. The number of samples in the training and test set are given in Table 1. Note that the number of test samples is larger than the number of training samples. This is due to the number of training images from the classes *Rock* and *Background* was intentionally limited in the training set to keep it balanced. The test set however is unbalanced with these two classes being overrepresented (see (Steiniger et al., 2023) for more details).

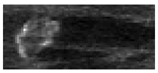
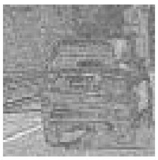
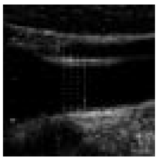
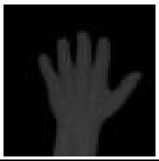
Dataset	Classes	Training snippets	Test snippets	Example
Sonar	5	129	1,486	
ImageNet	500	255,772	75,000	
ImageNet (grayscale)	500	255,772	75,000	
Malo	3	776	98	
Fetal	6	7,129	5,271	
Ships&Boats	2	812	49	
SAR	2	62,981	20,000	
Hand X-ray	6	40,637	10,160	
Synthetic	3	408	103	

Table 1. Datasets investigated for pre-training.

One of the most common datasets to pre-train deep learning models for the classification task is ImageNet. However, ImageNet contains optical RGB images with a higher resolution and more details than sonar images and thus is expected to have a large domain gap to the sidescan sonar dataset. During pre-training with ImageNet the network learns features based on color which are meaningless for the target task of classifying grayscaled sonar images. Thus, one criterion for selecting the datasets in this work was that the images should be grayscale to ensure a small domain gap. Typical sensors whose images fulfill this requirement are ultrasonic transducers, synthetic aperture radar (SAR) or X-ray. Another aspect for the selection was the number of training samples, since a dataset used for pre-training should contain more samples than the target dataset. We used the data sharing platforms Kaggle and Roboflow to search for open source datasets and selected the following ones: Malo (asd, 2023), Fetal (Burgos-Artizzu et al., 2020), Ship&Boats (Ng, 2023), SAR (Wang et al., 2019) and Hand X-ray (RF Projects, 2023). Originally, the SAR dataset contains ships to be detected. To use it for a classification task, we extract the ships and additional background snippets at random positions. This results in a binary classification dataset. In addition to these datasets, we manually modified the images from ImageNet to be grayscale. Table 1 gives an overview about the selected datasets regarding the number of classes and number of samples.

We also setup a simulation using a CAD software where we randomly placed models of ships, tires and rocks in a scene to generate a dataset of synthetic sonar images. Within the generated CAD environment, the essential features and components are: floor, light, and objects. The floor is an extruded planar surface which extends in a rectangular form. It includes different surface elevations along its area and in some other scenarios it also contained a ripple pattern using a simple $\sin()$ function. For better resemblance with real sonar images, an additional sand texture was set to the planar surface. In a second step, for enhancing the environment to a more realistic scene, two light sources were placed on the 3D assembly. By alternating its location and angle a wide range of possibilities in image diversity was achieved. Finally, the objects of interest, e.g., ships, boats and tires, were downloaded from a free CAD source. Only the rock class was manually created implementing a free-form option from the same CAD software.

3. MEASURED DOMAIN GAP

To quantify the domain gap between the pre-training datasets and the sidescan sonar dataset we first map the images to a lower dimensional space, as shown in Figure 2. Afterwards, the Euclidean distance between the center points of the datasets in this space is calculated. Mensink et al. propose to use a backbone CNN which was pre-trained on ImageNet to extract features of images from the individual dataset and calculate the distance between these feature vectors (Mensink et al., 2022). However, since ImageNet is one dataset to be investigated we disregard this approach. In this work, t-SNE as well as a VAE are used for the purpose of dimensionality reduction.

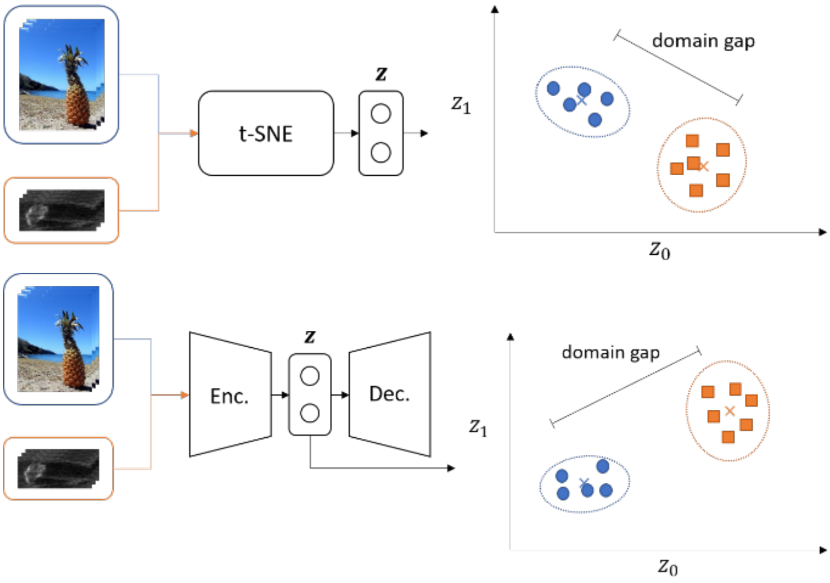
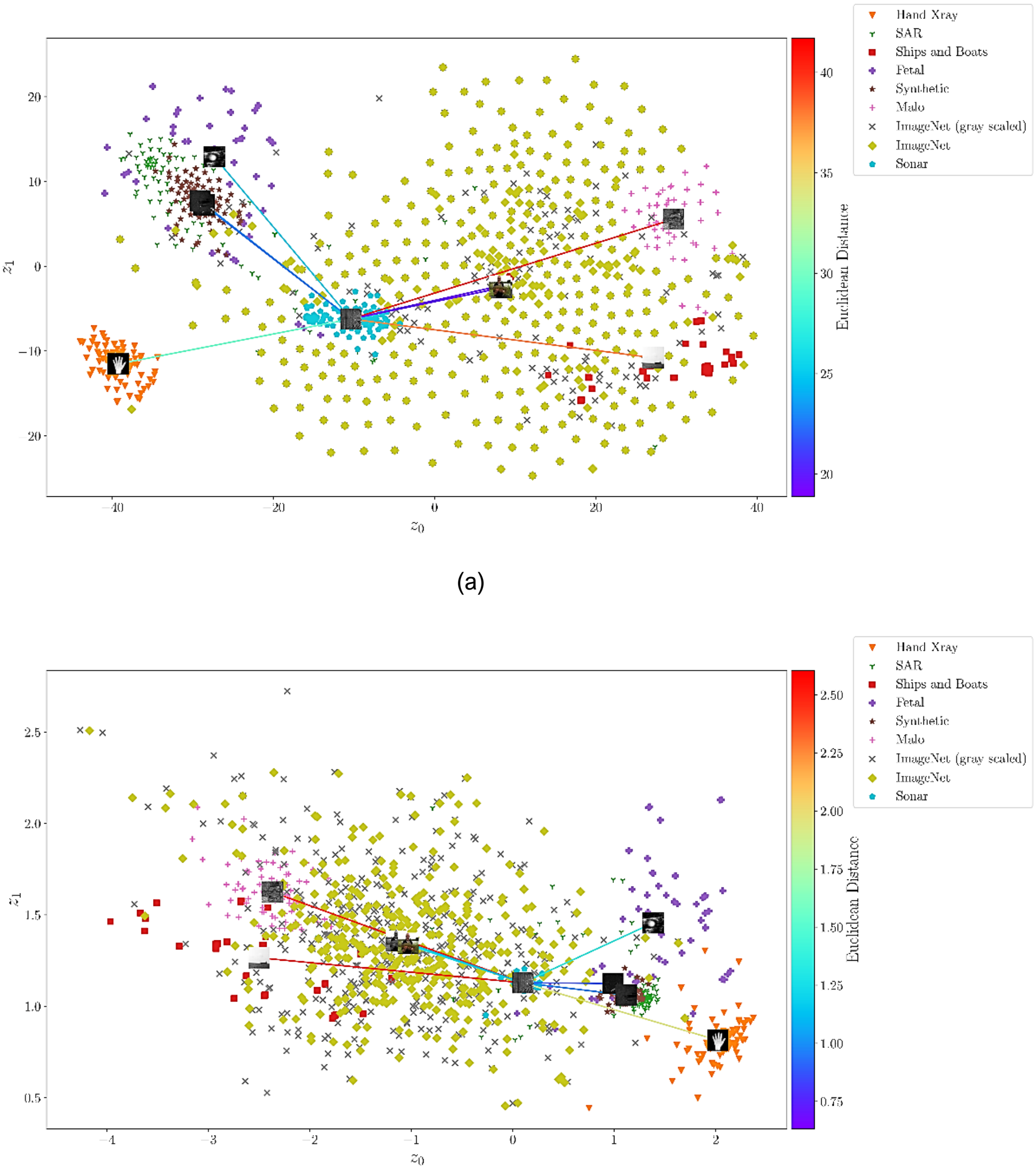


Figure 2: Measuring the domain gap using the t-SNE (top) and VAE (bottom) representation.

The t-SNE method maps the images to a lower dimensional space where images that are similar to each other should lie close to each other in the embedding. To achieve this, the Kullback-Leibler divergence between the distributions of the similarities of the images and of the similarities of the embedded points is minimized. The initial embeddings are determined based on a principle component analysis. In this work and in accordance with common practice we set the embedding space to be two-dimensional and run the optimization for 500 iterations. With the VAE, an encoder network maps the input images to a lower dimensional latent vector z and a decoder network learns to reconstruct the image from this. We design the VAE to have a latent vector of size two in order to obtain a two-dimensional representation of the images. The encoder consists of two convolutional layers with 32 and 64 kernels of size 3×3, respectively. The decoder reconstructs the image through a fully connected layer with 16,384 neurons and three transposed convolutional layers with 64, 32 and 1 kernel of size 3×3. We train the VAE for three epochs with the Adam optimizer in its standard configuration and a batch size of 128.

Because the input size of the classification networks has to be the same for all datasets and the RGB images from ImageNet have three color channels while all other datasets contain grayscale images with only one color channel, we repeat all grayscale images in the remaining two color channels. Figure 3 (a) and (b) show the distribution of the investigated datasets in the t-SNE and VAE embedding space. The center point of each dataset is calculated as the mean embeddings. Table 2 lists the Euclidean distances between the sonar and the different source datasets. In the VAE embedding space the SAR dataset has the shortest distance to the sonar image dataset. Surprisingly, although sonar images and optical images are fundamentally different, the distance between ImageNet and the sonar dataset measured in the t-SNE embedding is the smallest compared to the other datasets. This is partly due to the large spread of the ImageNet samples in the embedding space which is also present in the VAE embedding space. Comparing the distances for the original and the grayscaled version of ImageNet only a slight change is visible with the original one even having a smaller distance in the VAE representation. Intuitively, the grayscaled images are more similar to sonar images and should lead a smaller domain gap. This shows that both methods might not be optimal to measure the domain gap and further research needs to be done to quantify the domain gap between datasets.



(a)

(b)

Figure 3. Comparison of the domain gap. (a) Based on t-SNE. (b) Based on VAE.

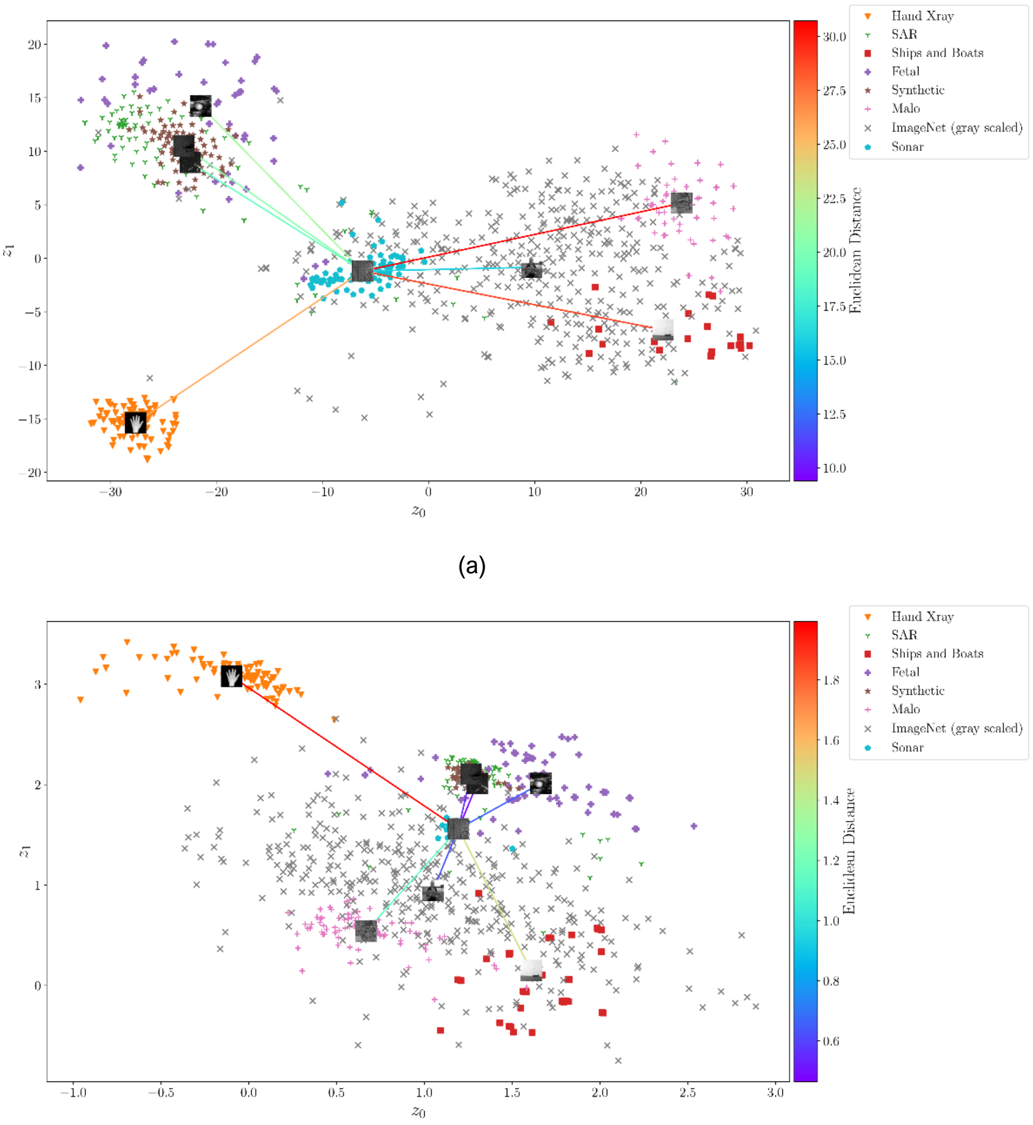
Dataset	Distance based on	
	t-SNE	VAE
ImageNet	**18.880**	1.150
ImageNet (grayscale)	19.029	1.266
Malo	41.713	2.516
Fetal	25.542	1.324
Ships&Boats	37.788	2.605
SAR	22.647	**0.888**
Hand X-ray	29.334	1.946
Synthetic	23.321	1.020

Table 2. Measured distances between sonar and pre-training dataset with the smallest distance indicated in bold.

Additionally, Figure 4 (a) and (b) show the distance plots for the grayscale images when the input to the network only consists of one channel. Thus, the original ImageNet is excluded. The corresponding distance measures are listed in Table 3. It can be seen that this modification has only minor influence on the distribution in the t-SNE representation. For the VAE the effect is slightly stronger. However, the general distribution of the datasets stays the same. In the three channels as well as the one channel case the datasets SAR, Synthetic and Fetal are grouped in the same area.

Dataset	Distance based on	
	t-SNE	VAE
ImageNet (grayscale)	**16.021**	0.732
Malo	31.045	1.163
Fetal	22.190	0.630
Ships&Boats	28.534	1.500
SAR	19.261	**0.437**
Hand X-ray	25.876	1.874
Synthetic	20.490	0.550

Table 3. Measured distances between sonar and pre-training dataset for one channel inputs with the smallest distance indicated in bold.



(a)

(b)

Figure 4. Comparison of the domain gap with one channel input. (a) Based on t-SNE. (b) Based on VAE.

4. NETWORK ARCHITECTURE AND TRAINING

We investigate the effect that the different pre-training datasets have on the classification performance of a broad variety of neural networks. This includes a shallow CNN used in our previous work (Steiniger et al., 2023), common CNNs for image classification as well as a transformer-based architecture. Our shallow CNN consists of three convolutional layers with 8, 16 and 32 kernels of size 3×3. The output of a convolutional layer is passed through a ReLU activation function, batch normalization and 2×2 max pooling. Features from the last convolutional layer are compressed using a fully connected layer with 100 neurons prior to the final output layer. Depending on the pre-training dataset the number of neurons in the output fully connected layer matches the number of classes for each dataset, e.g., the CNN trained on the Malo dataset has three output neurons. Dropout is added before both of the fully connected layers. All input images are scaled to match the input size of the network, which is 64×64 pixel. Furthermore, we use VGG-16 (Simonyan and Zisserman, 2015), ResNet-18, ResNet-101 (He et al., 2016), MobileNetv3 (Howard et al., 2019) and ViT-B/16 (Dosovitskiy et al., 2021) in our experiments. We selected these models based on their usage in general computer vision tasks and sonar image classification. In addition, considering different types of models, e.g., VGG-16 as a shallower network, MobileNetv3 as an edge-computing model and ViT-B/16 as a transformer-based model, allows to link our findings to properties of the network architecture.

We train all five standard models with their original input size. In order to adapt the models for the specific source datasets we replace their classification head with the classification head architecture of the custom CNN, i.e., a dropout layer followed by a fully connected layer with 100 neurons and ReLU activation function and the output layer. In the following experiments a model is first pre-trained on one of the source datasets and afterwards fine-tuned on the sidescan sonar image dataset. When pre-trained on datasets other than ImageNet, we consider two setups. In the first case the models are initialized with weights gained from training on ImageNet. This can be considered as an additional pre-training prior to the experiments and as a way to close the domain gap while at the same time benefiting from a large general pre-training dataset. Note that we were not able to train the custom CNN on ImageNet since its capacity is too low. In the second case we randomly initialize all weights. In all cases, pre-training of the standard networks is done for 20 epochs using the Adam optimizer. During the first 10 epochs only the adapted classification head is trained with a learning rate of 0.0001, while the weights of the remaining layers are kept fixed. Afterwards all layers are trained for the final 10 epochs with a learning rate of 0.000001. The custom CNN is trained for 50 epochs using a learning rate of 0.0001. As shown in (Gutstein et al., 2022) an optimal performance on the source dataset is not necessary to achieve a good transfer learning result. Thus, we did not focus on optimizing the performance of the models after pre-training. For fine-tuning on the sonar dataset, the output layer is adapted to match the sonar image classification task with five classes, i.e., the output fully connected layer now has five neurons. All networks are fine-tuned for 20 epochs with the Adam optimizer and a learning rate of 0.00001 which is reduced by a factor of 0.1 after 10 epochs. We experimented with different learning rates and found the reported configuration to work best. Additionally, we train all models on the sonar dataset without pre-training. This serves as a baseline, to investigate if pre-training is beneficial for sonar image classification.

5. CLASSIFICATION RESULTS

To account for the unbalanced test dataset, the classification performance of all models is assessed using the balanced accuracy

$$ACC_{bal} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} \mathrm{recall}(c) \qquad (1)$$

and the macro F1-score

$$F1_{macro} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} F1(c) = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} \frac{2\cdot\mathrm{precision}(c)\cdot\mathrm{recall}(c)}{\mathrm{precision}(c)+\mathrm{recall}(c)}, \qquad (2)$$

with $\mathcal{C} = \{$*Tire*, *Rock*, *Cylinder, Wreck, Background*$\}$. Figure 5 shows the baseline performance of the models without pre-training on the source datasets. For the standard models both cases, with and without ImageNet weight initialization, are shown. All standard models, except MobileNetv3 for the macro F1-score, show a better performance than the custom build CNN, regardless of their initialization method. ResNet-101 with random initialization is the best performing model with balanced accuracy of 0.394 and macro F1-score of 0.371. When initialized with ImageNet weights ResNet-101 also shows the best overall performance with a balanced accuracy of 0.350 and macro F1-score of 0.271. In the following experiments both will serve as the baseline performance for the individual initialization methods. ViT-B/16 is the only model that benefits from using ImageNet weights. One reason for this could be that the Transformer can learn more general features from ImageNet than the CNNs. However, even with ImageNet weights the performance of ResNet-101 is still better than the one of ViT-B/16. Transformer benefit mostly from very large datasets and the available sonar data could be too few to exploit the full capacity of the ViT-B/16.

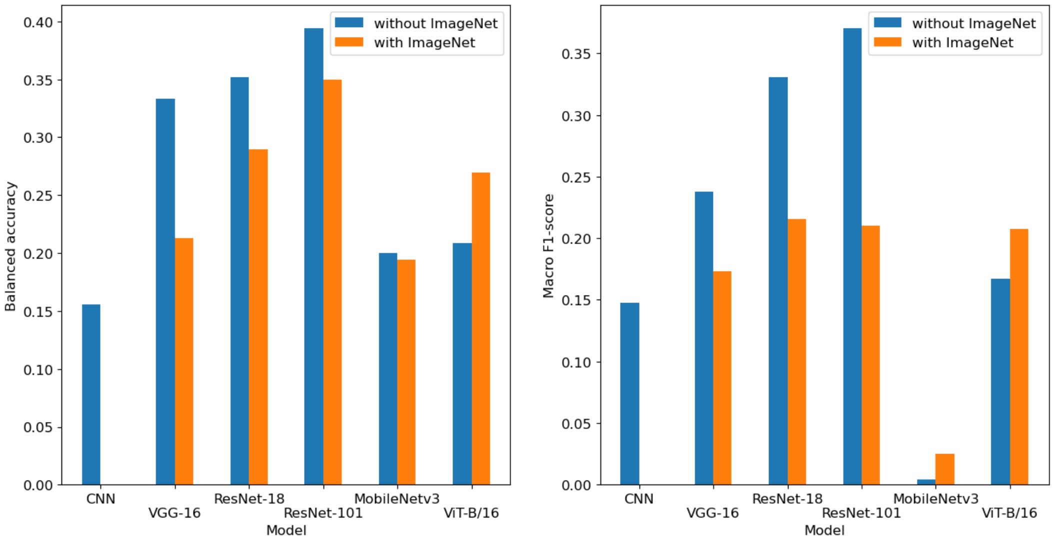


Figure 5. Performance of the models without pre-training on the source datasets. Orange bars show the results for standard models whose weights are initialized using ImageNet.

After investigating the baseline performance without pre-training, Figure 6 shows the classification performance of the individual models pre-trained on the seven source datasets and fine-tuned on the sonar dataset. The models were initialized with ImageNet weights, which is our first experimental setup for pre-training. Note that since the shallow CNN could not be trained on ImageNet it is not investigated in this initialization case. A green dot in Figure 6 indicates a model-dataset configuration which improves in the respective metric over the baseline ResNet-101 with ImageNet weight initialization. Interestingly, no configuration improves over the previously best performing baseline ResNet-101 with random initialization. One reason for this could be that the pre-training phase on the individual datasets is not sufficient to overwrite features that were learned from the ImageNet initialization but are not useful for sonar image classification. A red square indicates an improvement by pre-training this model on the specific dataset compared to directly training it on the sonar dataset. Especially the VGG-16 and MobileNetv3 benefit from pre-training. Using the grayscaled ImageNet dataset for pre-training leads to a worse performance for all models except MobileNetv3. The ViT-B/16 shows an improved macro F1-score for all datasets except grayscaled ImageNet and Ships&Boats. Figure 6 also shows that pre-training mostly improves the macro F1-score but not the balanced accuracy. In contrast to the balanced accuracy which only considers the recall, the macro F1-score also take the precision and thus the number of false-positive errors into account. For the unbalanced test dataset used in this work small improvements in the false-positive error of the underrepresented class have a relatively large effect on the precision and thus on the macro-averaged F1-score. At the same time, the effect of the corresponding improvement in false-negative error in the overrepresented class on the recall and subsequently on the balanced accuracy is minor in case of a large number of true-positives.

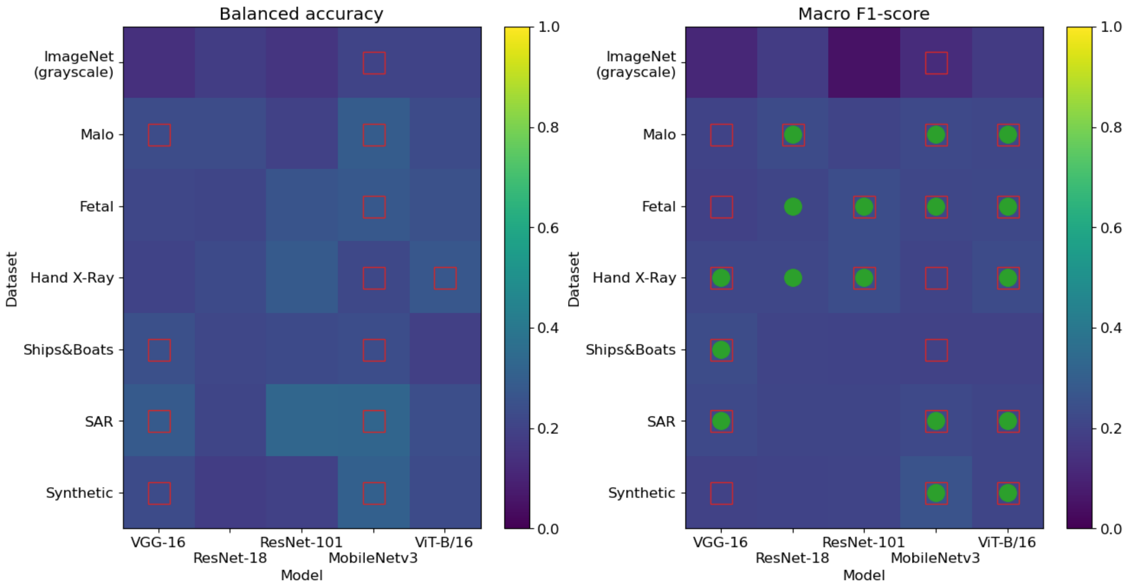


Figure 6. Performance after transfer-learning. Models were initialized with ImageNet weights. Green dots indicate an improvement over the baseline with ImageNet weights. Red squares indicate improvement compared to results without pre-training.

Figure 7 displays the classification performance of the models with random weight initialization when pre-trained on the individual source dataset. Note that the shallow CNN was not pre-trained on grayscaled ImageNet. Compared to the previous case where the models were initialized with ImageNet weights it can be seen that less configurations surpass the baseline in terms of the macro F1-score. However, four model-dataset configurations improve the baseline balanced accuracy with ResNet-18 pre-trained on the Fetal dataset even exceeding the best baseline performance of ResNet-101 directly trained on the sonar data with a balanced accuracy of 0.425. The shallow CNN pre-trained on Ships&Boats and ResNet-101 pre-trained on the synthetic dataset are the only configurations which improve accuracy as well as macro F1-score. For the case of random weight initialization pre-training is beneficial for the shallow CNN, MobileNetv3 and the ViT-B/16. However, both experiments have shown that no dataset can clearly be preferred over the others for pre-training the models. In addition, due to the small training dataset the achieved performance of all models could still be improved. Besides enlarging the dataset with more sonar images, generating synthetic data using generative deep learning models is a promising approach (Sanford et al., 2024).

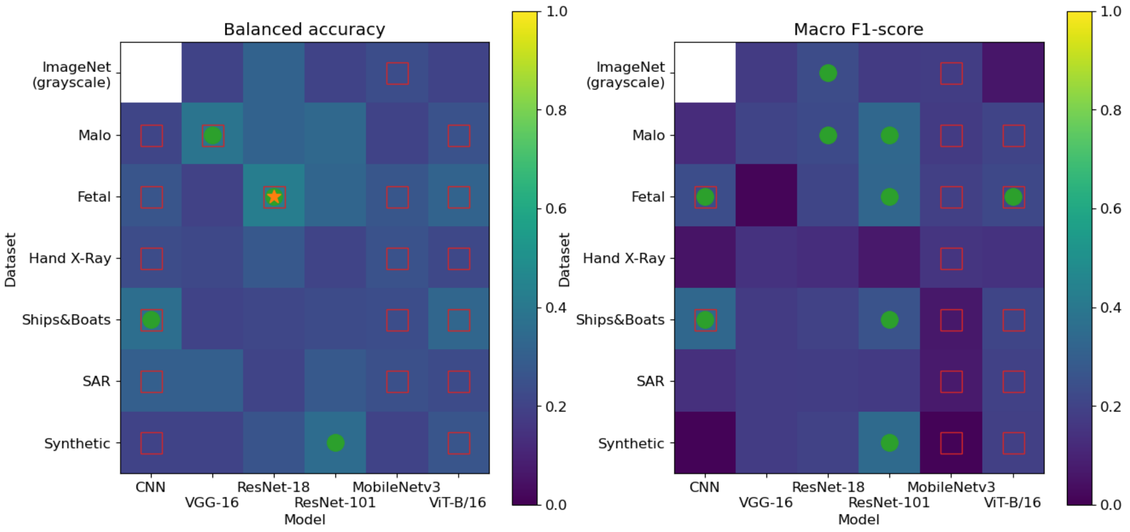
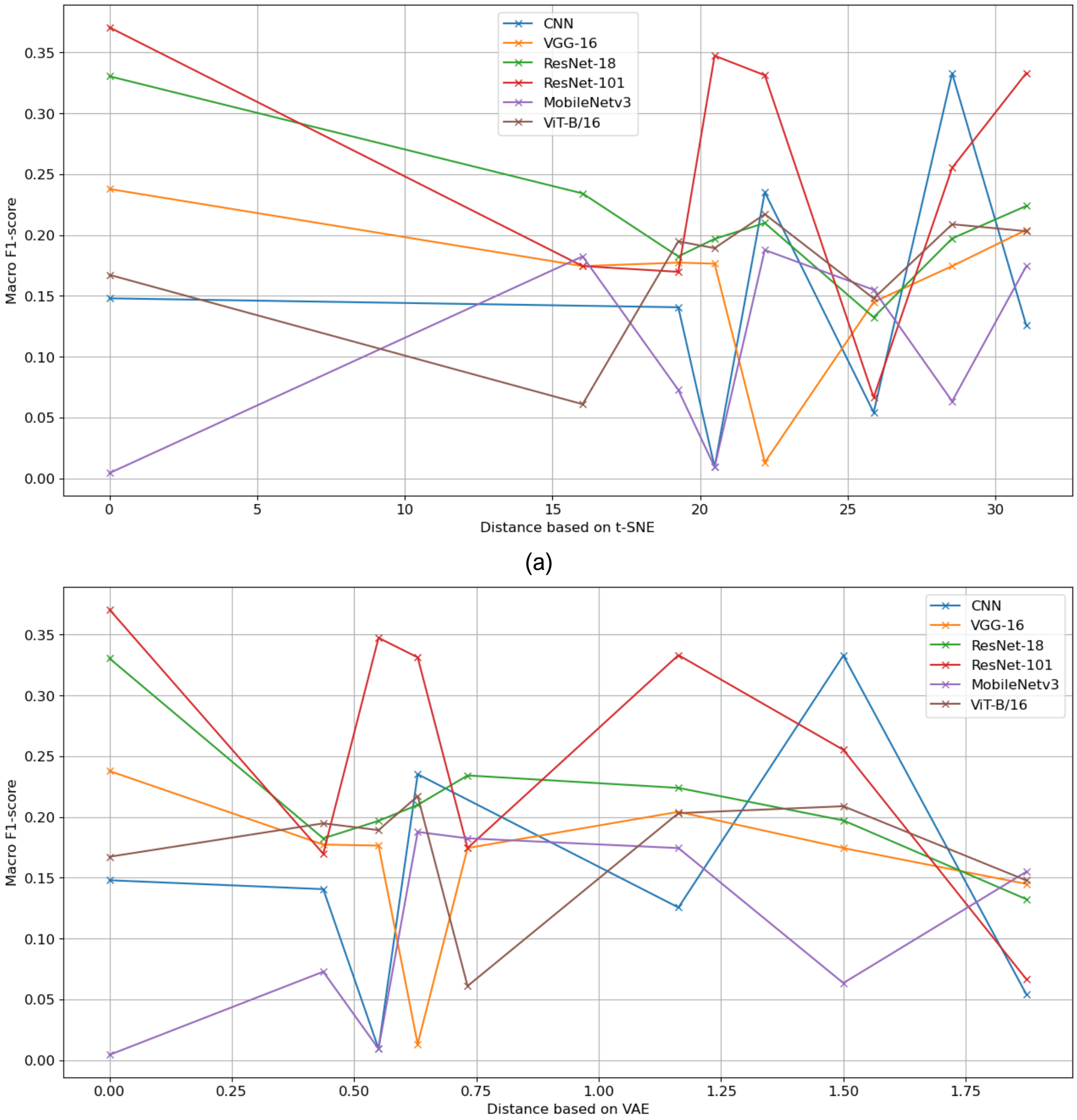


Figure 7. Performance with random initialization. Orange star indicates an improvement over the baseline without ImageNet and green dots an improvement over the baseline with ImageNet weights. Red squares indicate improvement compared results without pre-training.



(a)

(b)

Figure 8. Classification performance of the deep learning models vs. domain gap between pre-training and sonar dataset. (a) Based on t-SNE. (b) Based on VAE.

To directly link the classification performance and the measured domain gap, Figure 8 (a) and (b) plot the macro F1-score against the Euclidean distance in the t-SNE and VAE representation, respectively. Both figures show that the performance slightly drops with increasing distance between the source and the sonar dataset. Note however, that the distance not necessarily reflect the intuitive domain gap, since for example ImageNet shows the smallest distance in the t-SNE representation while expected to have the largest domain gap. Additionally, another important aspect when pre-training a deep learning model is the number of training samples. Comparing the classification performance with the size of the datasets given in Table 1 it can be seen that Fetal, which improves the performance of many of the considered models, has a relatively large domain gap but contains in total 7,129 training images and between 353 and 2,601 samples per class, which makes it one of the larger datasets studied in this work. This indicates that a small domain gap itself is not the only requirement for a good transfer-learning result. The source dataset also has to be sufficiently large.

6. CONCLUSION

This work has presented a study on different pre-training datasets and deep learning models for the classification of sonar images. Since the number of sonar snippets is limited, pre-training can be a beneficial way to learn relevant features in a first training step. However, the domain gap to the sonar image dataset should be small. Our analysis shows that distance measured in the t-SNE and VAE embedding space does not match the intuitive domain gap. Additional work needs to be done in order to measure the domain gap more convincingly. Furthermore, pre-training using the Fetal dataset led to an improvement for four of the six investigated models in terms of balanced accuracy. However, no general decision about which dataset should be used for pre-training can be made because no dataset significantly improved the balanced accuracy and macro F1-score of all models.

CONFLICT OF INTEREST

Authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Burgos-Artizzu, X.P. et al. (2020) FETAL_PLANES_DB: Common Maternal-Fetal Ultrasound Images. Available at: https://doi.org/10.5281/zenodo.3904280.

Deng, J. et al. (2009) ‘ImageNet: A large-scale hierarchical image database’, 2009 IEEE Conference on Computer Vision and Pattern Recognition. Available at: https://doi.org/10.1109/CVPR.2009.5206848.

Dosovitskiy, A. et al. (2021) ‘An image is worth 16x16 words: Transformers for image recognition at scale’, 9th International Conference on Learning Representations. Available at: https://doi.org/10.48550/arXiv.2010.11929.

Gutstein, S. et al. (2022) ‘Does optimal source task performance imply optimal pre-training for a target task?’, CoRR. Available at: https://doi.org/10.48550/arXiv.2106.11174.

He, K. et al. (2016) ‘Deep residual learning for image recognition’, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Available at: https://doi.org/10.1109/CVPR.2016.90.

Howard, A. et al. (2019) ‘Searching for MobileNetV3’, 2019 IEEE/CVF International Conference on Computer Vision. Available at: https://doi.org/10.1109/ICCV.2019.00140.

Mensink, T. et al. (2022) ‘Factors of influence for transfer learning across diverse appearance domains and task types’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12). Available at: https://doi.org/10.1109/TPAMI.2021.3129870.

Ng, C.W. (2023) Ship Detection Test Dataset. Available at: https://universe.roboflow.com/chee-wee-ng-efupv/ships-detection-test (Accessed: 26 June 2024).

Phung, S.L. et al. (2019) ‘Mine-like object sensing in sonar imagery with a compact deep learning architecture for scarce data’, 2019 Digital Image Computing: Techniques and Applications. Available at: https://doi.org/10.1109/DICTA47822.2019.8945982.

RF Projects (2023) X Ray Test Dataset. Available at: https://universe.roboflow.com/rf-projects/x-ray-test (Accessed: 26 June 2024).

Roboflow (2023) Malo Dataset. Available at: https://universe.roboflow.com/asd-vp0be/malo-vhhkp (Accessed: 26 June 2024).

Sanford, C. et al. (2024) ‘Fourier-domain wavefield rendering for rapid simulation of synthetic aperture sonar data’, IEEE Journal of Oceanic Engineering, 49(4). Available at: https://doi.org/10.1109/JOE.2024.3401968.

Sheffield, B.W. et al. (2024) ‘On vision transformers for classification tasks in side-scan sonar imagery’, International Conference on Underwater Acoustics 2024. Available at: https://doi.org/10.48550/arXiv.2409.12026.

Simonyan, K. and Zisserman, A. (2015) ‘Very deep convolutional networks for large-scale image recognition’, 3rd International Conference on Learning Representations. Available at: https://doi.org/10.48550/arXiv.1409.1556.

Steiniger, Y. et al. (2022) ‘Investigating the training of convolutional neural networks with limited sidescan sonar image datasets’, OCEANS 2022 MTS/IEEE – Hampton Roads. Available at: https://doi.org/10.1109/OCEANS47191.2022.9977255.

Steiniger, Y. et al. (2023) ‘Tackling data scarcity in sonar image classification with hybrid scattering neural networks’, OCEANS 2023 MTS/IEEE – Limerick. Available at: https://doi.org/10.1109/OCEANSLimerick52467.2023.10244457.

Urick, R.J. (2013) Principles of Underwater Sound. 3rd edn. Los Altos, CA: Peninsula Publishing.

Wang, Y. et al. (2019) ‘A SAR dataset of ship detection for deep learning under complex backgrounds’, Remote Sensing, 11(7). Available at: https://doi.org/10.3390/rs11070765.

Warakagoda, N. and Midtgaard, Ø. (2024) ‘Vision transformers for sonar image classification’, International Conference on Underwater Acoustics 2024. Available at: https://doi.org/10.25144/22252.

Williams, D.P. (2021) ‘On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery’, IEEE Journal of Oceanic Engineering, 46(1). Available at: https://doi.org/10.1109/JOE.2019.2963041.