CONVOLUTIONAL NEURAL NETWORK TRANSFER LEARNING FOR UNDERWATER OBJECT CLASSIFICATION

David P. Williams NATO STO CMRE, La Spezia, Italy

1 INTRODUCTION

Convolutional neural networks (CNNs) have recently achieved state-of-the-art performance on a wide range of image classification tasks[1-3]. But to do so, vast amounts of labeled data are required. Unfortunately, in remote-sensing applications, data collections can be time-consuming and prohibitively expensive. Moreover, when a new sensor is introduced, there is a desire to be able to still leverage historical data collected by similar predecessor systems. In general, it is not feasible to wait for the execution of numerous onerous data collections before being able to accurately assess the utility of the new system. For these reasons, we explore the idea of transfer learning with CNNs. Essentially, the objective is to demonstrate that data from multiple similar remote-sensing sensors can be collectively exploited to address related automatic target recognition (ATR) tasks and to ease training data requirements.

CNNs were previously developed for mine classification with synthetic aperture sonar (SAS) data collected by the MUSCLE autonomous underwater vehicle (AUV)[4-6]. Leveraging this work, we demonstrate the feasibility of two types of transfer learning for the task of underwater object classification: target-concept transfer and sensor transfer. Specifically, we modify the target concept of the networks, from mines to unexploded ordnance (UXO), so that the objective is to successfully discriminate UXO – rather than mines – from clutter. The second type of CNN transfer learning we demonstrate involves transfer between sensors: training a CNN using SAS data from one sensor, and adapting it to enable inference on SAS data from a different sensor, operating at a different frequency band. Experimental results on real, measured SAS imagery illustrate the feasibility of these forms of CNN transfer learning.

The remainder of this paper is organized as follows. The necessary background on CNNs is provided in Sec. 2. Results on target-concept transfer-learning are shown in Sec. 3, while results on sensor transfer-learning are shown in Sec. 4. Concluding remarks are given in Sec. 5.

2 CONVOLUTIONAL NEURAL NETWORKS

A CNN is a sophisticated classification algorithm whose power derives from its great representational capacity. The standard architecture of a CNN consists of alternating layers of convolution and pooling operations, followed by a fully-connected layer, and a final (fully-connected output) prediction layer. The output of one layer is the input to the subsequent layer, with this nested functional structure -- in conjunction with nonlinear activation functions -- enabling highly complex decision surfaces. The input to a CNN is an image, and the outputs are the probabilities of belonging to each class under consideration (here, targets and clutter). Training a CNN means learning the parameters of the filters (and bias terms). A schematic representation of this basic architecture is shown in Figure 1.

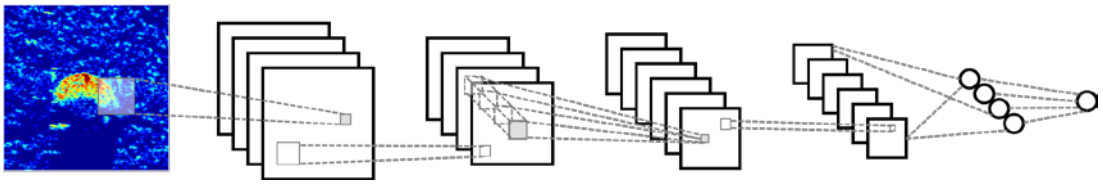

Figure 1. Schematic of basic CNN architecture consisting of an input image chip, a convolutional layer with 4 filters, a pooling layer, a convolutional layer with 6 filters, another pooling layer, a fully-connected layer, and the final class probability output.

For our application, the inputs to the initial layer are the SAS magnitude or phase "chips" of alarms flagged in the detection stage by the Mondrian detection algorithm[7] on larger SAS scene-level imagery (that typically spans 50 m x 110 m). The size of these input chip images are 267 pixels by 267 pixels, with a resolution of 1.5 cm in each dimension. The outputs of the final layer are the probabilities of a chip belonging to each class (target or clutter). Each convolutional layer and fully-connected layer uses a sigmoid activation function, while each pooling layer uses pure averaging rather than the commonly used max-pooling approach. Each convolutional layer is associated with a fixed number of filters (i.e., kernels) of predefined size.

The training process of the deep network learns the parameters of the model, which for the convolutional layers are the filters and associated bias terms. (There are no parameters associated with the pooling layers.) The model seeks to minimize the standard classification error on the training data under consideration. At each training iteration, the model parameters are updated by batch gradient descent. Because there can be thousands or even millions of free model parameters to be learned, it is necessary to have an extremely large set of training data to avoid overfitting. In turn, training a CNN "from scratch" can take considerable time, even with high-throughput computational resources like graphics processing units (GPUs).

In this work, we leverage seven CNNs previously trained for a mine recognition task by using a large database of SAS imagery collected during eight sea experiments in diverse locations[4-6]. Each CNN is distinguished by the number of convolutional layers, the numbers and sizes (in pixels) of the filters, the pooling factors employed, and the type of input data (e.g., magnitude image or phase image) assumed. The basic architectures of these CNNs are summarized in Table I, where the number of convolutional layers employed is equal to the number of elements in a given column. The number of free parameters to be learned in each CNN is on the order of 10^4, which is relatively small for CNNs. For context, the popular VGG-net[8] has on the order of 10^8 parameters.

TABLE I
CNN ARCHITECTURES

CNN Name	Input Data	Numbers of Filters	Sizes of Filters	Pooling Factors
A	mag.	8, 10, 12	16, 8, 5	4, 4, 2
B	mag.	8, 10, 12	8, 6, 6	4, 4, 2
C	mag.	2, 3, 4, 5	18, 16, 14, 12	2, 2, 2, 2
D	mag.	4, 6, 8, 10, 12	8, 7, 7, 5, 3	2, 2, 2, 2, 2
E	mag.	10, 10, 10, 10, 10	4, 3, 4, 4, 3	2, 2, 2, 2, 2
F	mag.	10, 10, 10, 10, 10, 10	2, 2, 3, 3, 2, 2	2, 2, 2, 2, 2, 2
G	phase	8, 10, 12	8, 6, 6	4, 4, 2

3 TARGET-CONCEPT TRANSFER-LEARNING

We first explore the idea of target-concept transfer-learning by altering the classes of objects considered targets and clutter. The objective is to refine already-trained CNNs so that they properly discriminate the new object classes without the burden of training the networks "from scratch."

3.1 Experimental Set-Up

Previously, SAS data collected by the MUSCLE AUV during eight sea experiments had been used to train seven CNNs in which the target class consisted of mine-like object shapes including cylinders, truncated cones, wedges, and other man-made objects. All other alarms were assigned to the clutter class. The center frequency of the MUSCLE SAS is 300 kHz, and the bandwidth is 60 kHz.

For these transfer-learning experiments, MUSCLE SAS data from three different sea experiments (not used in the earlier training process) were considered as test data. To effect a transfer-learning scenario, the target class was modified to consist of only cylindrical objects (as a surrogate for UXO, which typically takes this shape). All other objects (including the objects previously treated as targets) were considered to belong to the clutter class.

With the training data "relabeled" in this manner, CNN refinement was undertaken. The previously-trained CNN parameters were updated via batch gradient descent, using a batch of 32 data points per class, and a fixed learning rate of 1.0 in conjunction with a misclassification error loss-function. At each iteration, the data used for each class's batch was sampled from a larger pool of 38 randomly selected data points, with the sampling bias made to favor (i.e., choose) more difficult cases as quantified by the Mondrian detection scores. To augment the data set and improve robustness, each alarm's image was randomly reflected about an axis in the range direction, and randomly translated as well. Each input data point for the CNNs corresponds to a 4 m x 4 m SAS image chip, as in Figure 1.

Transfer-learning refinement was executed for 2000 epochs, where one epoch is defined to correspond to an update from a single batch, not a full pass through the entire data set. For the first 1000 epochs of re-training, the clutter class data points were constrained to be drawn from the subset of objects that were treated as targets in the original training, but as clutter in the refinement phase. For the second 1000 epochs, the clutter class data points were allowed to be drawn from all alarms assigned a clutter label for the re-training. This procedure was undertaken to examine whether it was sufficient to focus the refinement on "un-learning" the newly labeled clutter cases, or if considering the full set of clutter cases was necessary.

3.2 Results

The results of the target-concept transfer-learning experiments are summarized in Tables II-IV in terms of the area under the receiver operating characteristic (ROC) curve (AUC). Specifically, the AUC is shown for each of the seven CNNs considered, for each of the three test data sets, at epoch 0 (i.e., before any refinement had transpired), at epoch 1000, and at epoch 2000.

As can be seen from the tables, before refinement has commenced, performance is quite poor because the CNNs had been trained to treat certain objects as targets, which were then considered as clutter during the test phase. Upon refining the CNN parameters with the new labeling rubric, performance improved dramatically. It can be noted that considering the full set of clutter cases was needed to achieve the best performance, indicating that "un-learning" the altered-label cases alone was insufficient. It can also be seen from the tables that CNN G, which uses SAS *phase* imagery as input, was unable to improve performance with refinement, suggesting that the altered-label clutter objects were indistinguishable (from the target class) in this data representation.

TABLE II
TARGET-CONCEPT TRANSFER RESULTS ON GAM1 DATA SET

CNN Name	AUC at Epoch 0	Epoch 1000	Epoch 2000
A	0.633	0.829	0.851
B	0.467	0.788	0.840
C	0.398	0.753	0.732
D	0.479	0.842	0.837
E	0.496	0.780	0.748
F	0.554	0.836	0.791
G	0.619	0.519	0.586

TABLE III
TARGET-CONCEPT TRANSFER RESULTS ON ONM1 DATA SET

CNN Name	AUC at Epoch 0	Epoch 1000	Epoch 2000
A	0.774	0.737	0.898
B	0.799	0.782	0.939
C	0.694	0.870	0.990
D	0.738	0.844	0.915
E	0.797	0.674	0.932
F	0.709	0.773	0.904
G	0.645	0.488	0.645

TABLE IV
TARGET-CONCEPT TRANSFER RESULTS ON TJM1 DATA SET

CNN Name	AUC at Epoch 0	Epoch 1000	Epoch 2000
A	0.825	0.659	0.984
B	0.841	0.846	0.975
C	0.839	0.878	0.943
D	0.845	0.957	0.966
E	0.832	0.826	0.962
F	0.842	0.825	0.961
G	0.810	0.521	0.794

The evolution of the performance in terms of full ROC curves as a function of refinement epoch, for the six CNNs using magnitude imagery as input, is also shown in Figure 2 for the TJM1 test data set.

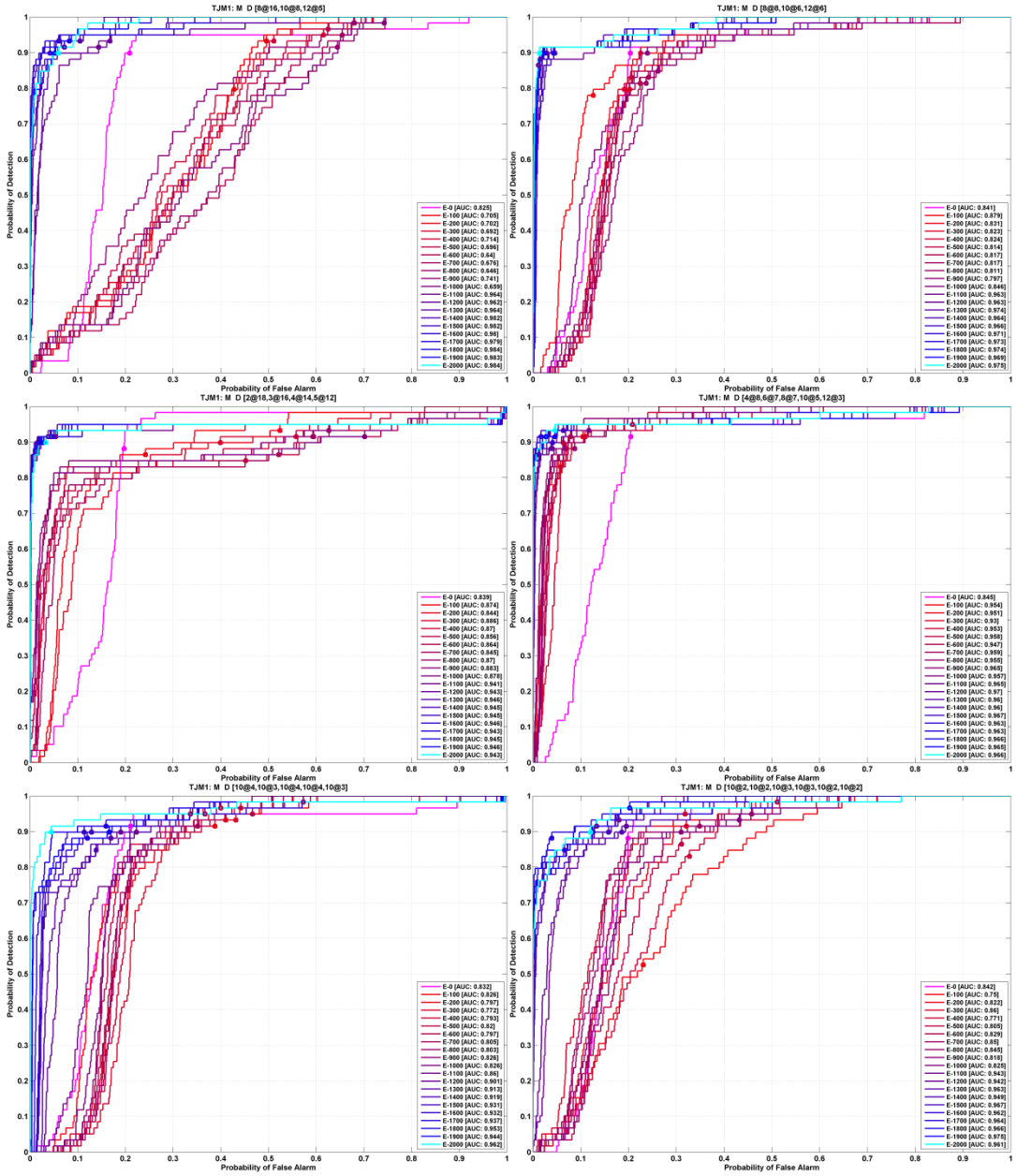

Figure 2. Performance on the TJM1 test data set of the six CNNs that use magnitude imagery. (Sub-figures, row-wise, left to right, show the results of CNN A through CNN F.) The initial performance before any refinement is shown in magenta; the final performance after 2000 epochs of refinement is shown in cyan. In between, the performance curves progress from red to blue with increasing epoch. (Zoom on the electronic version to read the legend and axes.)

4 SENSOR TRANSFER-LEARNING

Next we address the idea of sensor transfer-learning by considering data from a new sensor. The objective is to begin with CNNs that were trained used MUSCLE data, and refine them using a small amount of data from a different sensor, namely the SAS system on the SeaOtter Mk II AUV, so that the networks can properly classify test data from the new sensor. This approach would obviate the need for a large set of labeled training data from the new sensor, as would otherwise be required if training the networks "from scratch."

4.1 Experimental Set-Up

In this study, we use SAS data collected by the SeaOtter Mk II AUV in German waters in September 2016. The center frequency of the SeaOtter SAS is 150 kHz, and the bandwidth is 30 kHz. The data set considered here consisted of 476 SAS scene images, collectively spanning approximately 2.61 square kilometers of seabed.

The Mondrian detection algorithm was applied to the scene imagery, with this generating a set of 29208 candidate alarms. The survey data was collected in an area that is known to contain real UXO from World War II, but no proper ground truth was available. Therefore, we manually labeled all of the candidate alarms, assigning suspected UXO to the target class and all other alarms to the clutter class. (Although the resultant classification rates attained may be inaccurate, this approach will still fairly assess the potential of sensor transfer-learning.)

To form relatively balanced training and test data sets, the alarms generated from the port sonar were treated as training data, while the alarms generated from the starboard sonar were treated as test data. This division resulted in 65 and 73 targets in the training and test sets, respectively.

We again begin with the seven CNNs that had been trained using SAS data collected by the MUSCLE AUV during eight sea experiments, as in Sec. 3. But then transfer learning is effected by refining these CNNs with the training data set from the SeaOtter. This refinement used the same learning rate, batch size, and data augmentation techniques as in Sec. 3. The slightly lower resolution SeaOtter image chips were upsampled (via linear interpolation) to match the pixel size of the previous MUSCLE training data (i.e., 1.5 cm x 1.5 cm). The same image normalization procedure that had been applied to MUSCLE data was also applied to the SeaOtter data. Transfer-learning refinement was executed for 2000 epochs, where one epoch is again defined to correspond to an update from a single batch, not a full pass through the entire data set.

4.2 Results

The results of the sensor transfer-learning experiments are summarized in Table V in terms of the AUC. Specifically, the AUC is shown for each of the seven CNNs considered, as well as the ensemble, for the SeaOtter test data set, before any refinement had transpired (i.e., at epoch 0) and with refinement (at epoch 2000). The ensemble case uses the mean of the seven CNNs' predictions for an alarm as its final prediction.

Before refinement, the CNNs are still tailored to both the general characteristics of the MUSCLE sensor data and the specific target types (i.e., surrogate mine shapes) used for training. As a result, classification performance on the SeaOtter test data is relatively weak. As the CNNs are exposed to more of the SeaOtter training data during the refinement process, performance improves. This result can be observed in both Table V as well as Figure 3, which shows the progression of the full ROC curves as a function of refinement epoch. In fact, it can be seen that the refinement had more or less converged even after only 500 epochs.

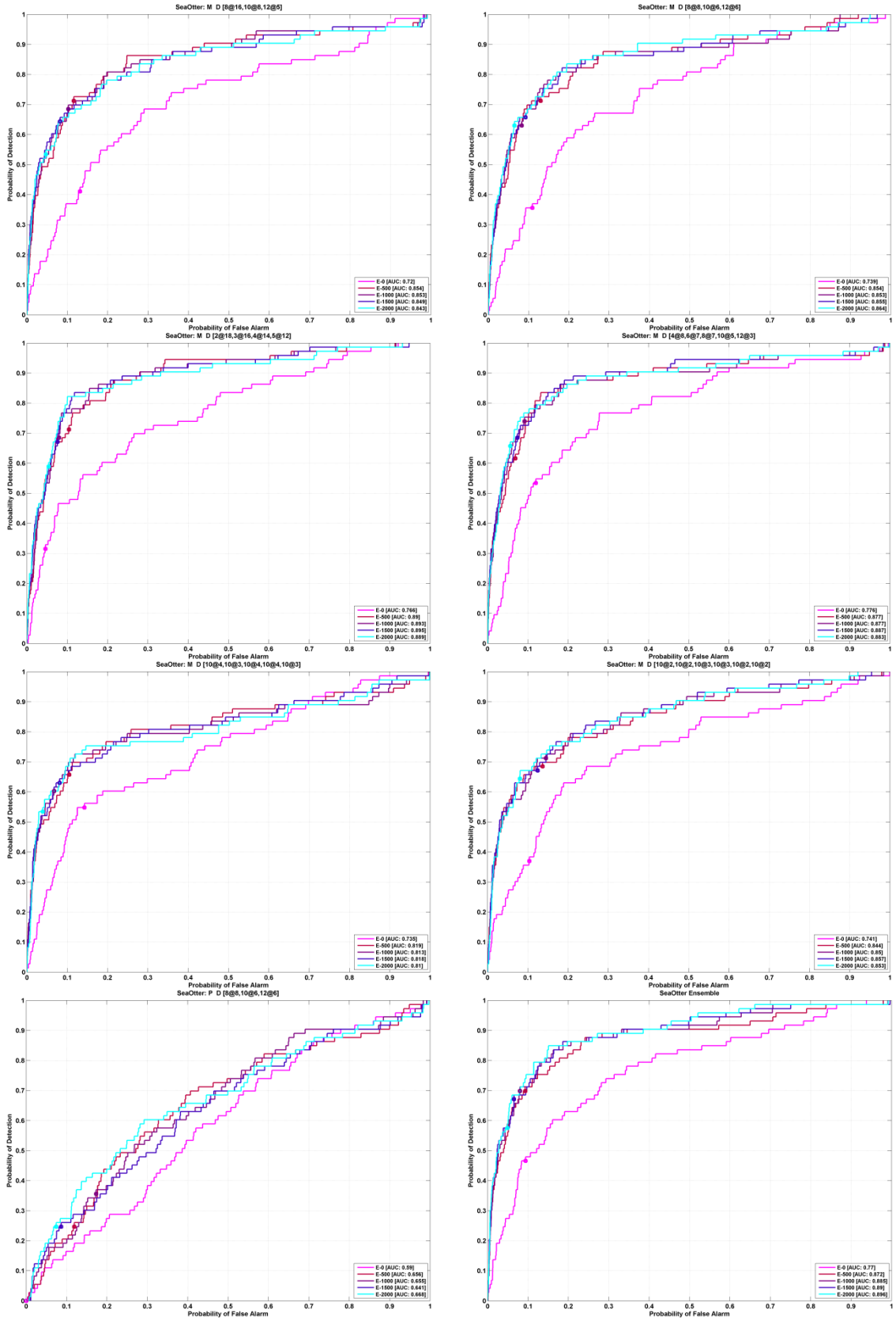

Figure 3. Performance of the seven CNNs, and the ensemble, on the SeaOtter test data set. (Sub-figures, row-wise, left to right, show the results of CNN A through CNN G, and then the ensemble.)

TABLE V
SENSOR TRANSFER RESULTS ON SEAOTTER DATA SET

CNN Name	AUC Without Refinement	AUC With Refinement
A	0.720	0.843
B	0.739	0.864
C	0.766	0.889
D	0.776	0.883
E	0.735	0.810
F	0.741	0.853
G	0.590	0.668
Ensemble (A-G)	0.770	0.896

Interestingly, relatively small amounts of SeaOtter training data are required to improve performance considerably. For example, only 65 examples of the new target class are available during the refinement process. This underscores the promise of the transfer-learning approach, and the minimal labeled-data requirements involved. Instead, one of the primary factors for success is that the test data and training data of the new sensor are drawn from the same underlying statistical distribution.

5 CONCLUSION

We demonstrated the feasibility of two types of transfer learning for the task of underwater object classification: target-concept transfer and sensor transfer. This transfer learning leveraged CNNs trained for a mine classification task with data from one SAS system, and performed parameter refinement using a small amount of data related to the new task. The success of the transfer learning suggests that CNNs developed on MUSCLE data can be exploited for use with data from other side-looking sonar systems that operate at different frequency bands and produce imagery at a different resolution; application to similar yet distinct classification tasks, such as a different target class, is also feasible. In turn, powerful classifiers can be developed relatively quickly with minimal labeled data, thereby reducing the amount of costly data surveys needed with the new sensor, as well as the computation time needed to train a robust classifier.

6 ACKNOWLEDGMENTS

The author would like to thank Holger Schmaljohann from WTD-71 for providing the SeaOtter data, and Isaac Gerg from PSU-ARL for assisting with the manual labeling of that data. This work was partially supported by the Strategic Environmental Research and Development Program (SERDP) and the NATO Allied Command Transformation (ACT).

7 REFERENCES

1. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification With Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
2. D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, “Mitosis Detection in Breast Cancer Histology Images With Deep Neural Networks,” in *Medical Image Computing and Computer-Assisted Intervention*, pp. 411–418. 2013.
3. D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the Game of Go With Deep Neural Networks and Tree Search,” *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
4. D. Williams, “Demystifying Deep Convolutional Networks for Sonar Image Classification,” *Proceedings of the Underwater Acoustics Conference*, Skiathos, Greece, June 2017.
5. D. Williams, “Underwater Target Classification in Synthetic Aperture Sonar Imagery Using Deep Convolutional Neural Networks,” *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)*, Cancún, Mexico, December 2016.
6. D. Williams, “Exploiting Phase Information in Synthetic Aperture Sonar Images for Target Classification,” *Proceedings of IEEE OCEANS 2018*, Kobe, Japan, May 2018.
7. D. Williams, “The Mondrian Detection Algorithm for Sonar Imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 2, pp. 1091-1102, February 2018.
8. K. Simonyan and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.