Exploiting Phase Information in Synthetic Aperture Sonar Images for Target Classification

David P. Williams
NATO STO
Centre for Maritime Research and Experimentation
La Spezia, Italy
david.williams@cmre.nato.int

***Abstract*—It is demonstrated that the phase information present in complex high-frequency synthetic aperture sonar (SAS) imagery can be exploited for successful object classification. That is, without using the amplitude content of the imagery, man-made targets can be discriminated from naturally occurring clutter. To exploit the information ostensibly hidden in the phase imagery, relatively simple convolutional neural networks (CNNs) are trained, "from scratch," on a large database of SAS phase images collected at sea. Inference is then performed on real SAS data collected at sea during five other surveys that span multiple geographical locations and a variety of seafloor types and conditions. These experimental results on the test data illustrate that the phase information alone can produce favorable object classification performance. To our knowledge, this work is the first to demonstrate this finding.**

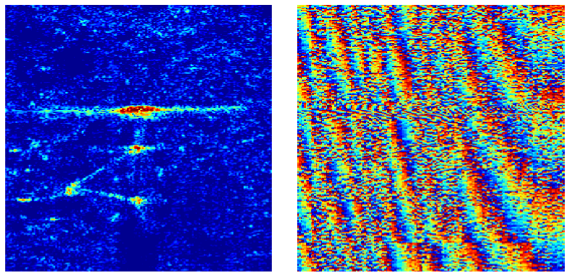
***Index Terms*—Classification, phase information, convolutional neural networks (CNNs), automatic target recognition (ATR), synthetic aperture sonar (SAS)**

I. Introduction

Synthetic aperture sonar (SAS) works by coherently summing received acoustic signals of overlapping elements in an array. Importantly, the resulting high-resolution SAS imagery is complex-valued. Typically this data is converted to a (real-valued) amplitude representation that is subsequently used for various signal processing and pattern recognition tasks, such as object classification. The phase information, which is related to the signal travel time, and in turn, the distance traveled, is usually discarded. In this work, we demonstrate that the phase data in SAS imagery contains useful information that can be exploited *on its own* for object classification tasks. This finding upends the conventional wisdom that the phase does not contain useful information for classification.

Complex SAS data is often manipulated to serve various purposes, but seldom is the phase information considered in isolation. (A notable exception is with bathymetric estimation via interferometry [1], where phase differences are exploited to obtain relative height information.) For example, transforming complex data into the Fourier domain enables efficient sub-band [2], [3] and sub-aperture processing [4], [5] and, with lower-frequency data, the formation of acoustic color representations [6]. But the general view of SAS phase imagery – as a signal with no worthwhile content – is an assumption we challenge.

An example SAS "mugshot" of an object – an endfire cylinder, with deployment chains attached – is shown in Fig. 1. Specifically, both the amplitude and phase images are shown. From visual inspection, it is obvious that the amplitude image contains useful information for classification. Less clear is whether the phase image also contains features or characteristics that can be exploited reliably. The objective of this work is to establish that phase imagery like this does indeed contain information. In order to demonstrate this, we rely on convolutional neural networks (CNNs) [7], which have the ability to automatically uncover useful clues for classification via its learned filters. This aspect of CNNs is particularly attractive because, as Fig. 1(b) suggests, it is challenging for a human to hand-craft salient features for extraction from phase imagery.



(a) Amplitude image (b) Phase image

Fig. 1. An object's SAS (a) amplitude image and the corresponding (b) phase image.

II. 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

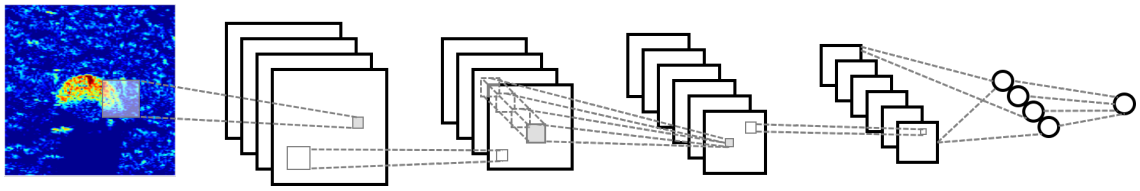


Fig. 2. Schematic of basic CNN architecture consisting of an input image (shown here as a SAS *amplitude* mugshot for illustrative purposes), a convolutional layer with 4 filters, a pooling layer, a convolutional layer with 6 filters, another pooling layer, a fully-connected layer, and the final class probability output.

TABLE I
CNN ARCHITECTURES

CNN Name	Numbers of Filters	Sizes of Filters	Pooling Factors
A	8, 10, 12	16, 8, 5	4, 4, 2
B	8, 10, 12	8, 6, 6	4, 4, 2
C	4, 6, 8, 10, 12	8, 7, 7, 5, 3	2, 2, 2, 2, 2

TABLE II
TEST DATA SET DETAILS

Data Set Code	Name of Sea Experiment	Survey Dates (months / year)	Survey Location	Survey Area (km^2)	Number of Targets	Number of Clutter
MAN2	MANEX	9-10 / 2014	Bonassola, Italy	38.9	375	77222
NSM1	NSMEX	5 / 2015	Ostend, Belgium	22.6	52	46832
TJM1	TJMEX	10 / 2015	Cartagena, Spain	55.5	357	43847
ONM1	ONMEX	9 / 2016	Hyères, France	43.8	71	23366
GAM1	GAMEX	3-4 / 2017	Patras, Greece	19.4	72	4058

with nonlinear activation functions – enabling highly complex decision surfaces. The input to a CNN is an image, as in Fig. 1, 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 Fig. 2.

For our application, the inputs to the initial layer are the SAS phase "mugshots" of alarms flagged in the detection stage by the Mondrian detection algorithm [8] on larger SAS scene-level imagery (that typically spans 50 m $\times$ 110 m). The size of these input mugshot images are 267 pixels by 267 pixels, with a resolution of 1.5 cm in each dimension. All pixel values are in $[0, 2\pi)$. The outputs of the final layer are the probabilities of a mugshot 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. In contrast to the computer vision community, where the size of the filters is usually only a few pixels wide, we use larger filters in order to permit the uncovering of richer, more meaningful characteristics in the data (that can hopefully be tied to physical phenomena).

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 a form of stochastic 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 many months, even with high-throughput computational resources like graphics processing units (GPUs).

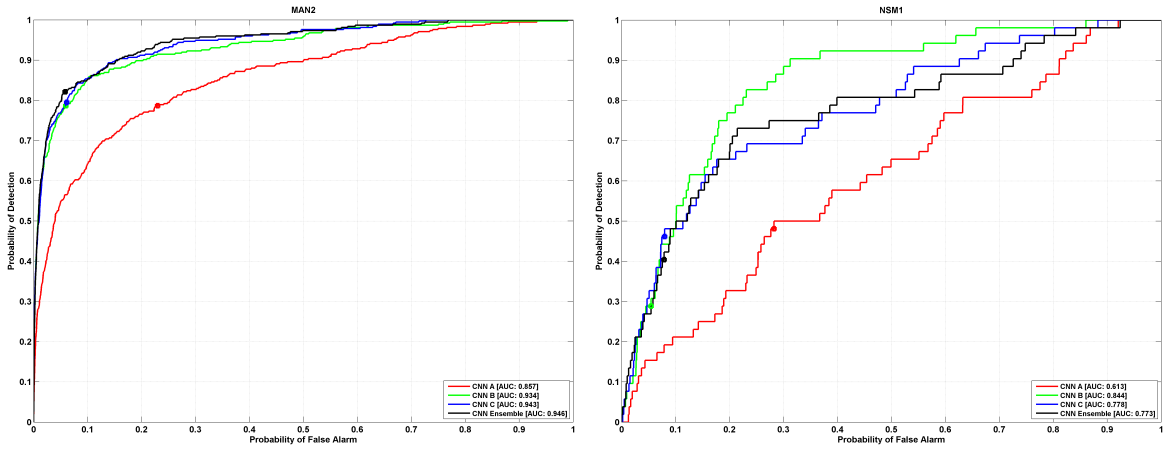
In this work, we develop three unique CNNs, distinguished by the number of convolutional layers, the numbers and sizes (in pixels) of the filters, and the pooling factors employed. The basic architectures of the CNNs designed are summarized in Table I, where the number of convolutional layers employed 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.

To train these CNNs, a large database of SAS phase imagery collected during eight sea experiments in diverse locations was used [9]. Testing was then performed using a disjoint set of SAS data from five other sea experiments. Basic details of these test sets are summarized in Table II.

III. EXPERIMENTAL RESULTS

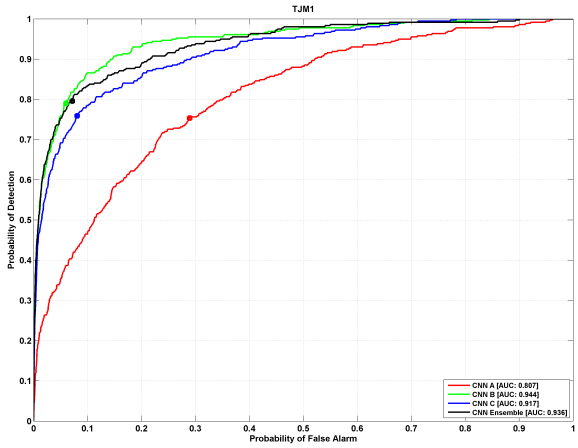
The results of making (class) predictions using the three CNNs for data from the five test sets are shown in Fig. 3, where it can be seen that the classification performance is well above the chance diagonal (in which every prediction is a random coin flip). This result provides strong evidence that there is indeed exploitable classification information contained in the phase images alone. The area under the receiver operating characteristic (ROC) curve (AUC) associated with Fig. 3 is also shown in Table III.

Interestingly, classification performance was markedly worse on the NSM1 data set. This data was collected in the North Sea where there were very strong currents. As a result, the sonar-equipped autonomous underwater vehicle (AUV) was rarely able to maintain an ideal linear trajectory during data collection, and SAS processing image-formation was more challenging. The *amplitude* imagery from this data set is often blurry, and shadows cast by objects are typically not well-defined. We hypothesize that this factor also causes the phase imagery to lack strong structure in the shadows, thereby eliminating exploitable classification clues.

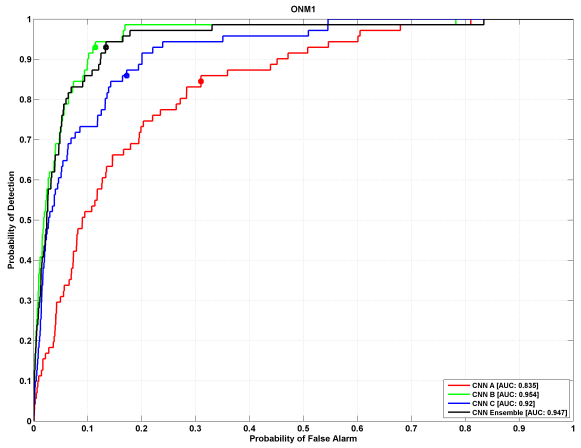


(a) MAN2

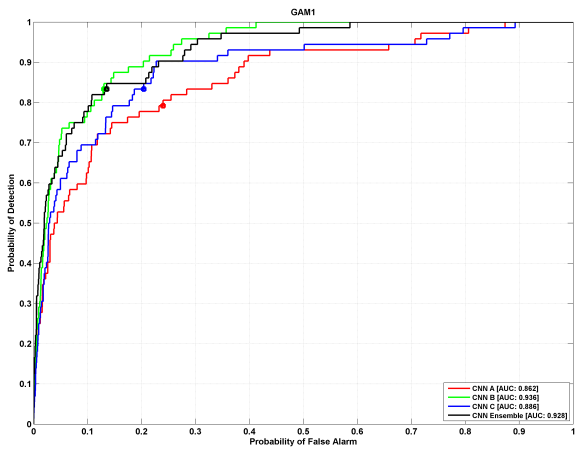
(b) NSM1



(c) TJM1



(d) ONM1



(e) GAM1

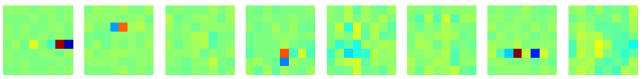
Fig. 3. Classification performance on five different test data sets, indicated by sea trial code, using three different CNNs and the ensemble. The operating point corresponding to a prediction threshold of 0.5 is marked, by a circle, on each curve.

IV. Analysis

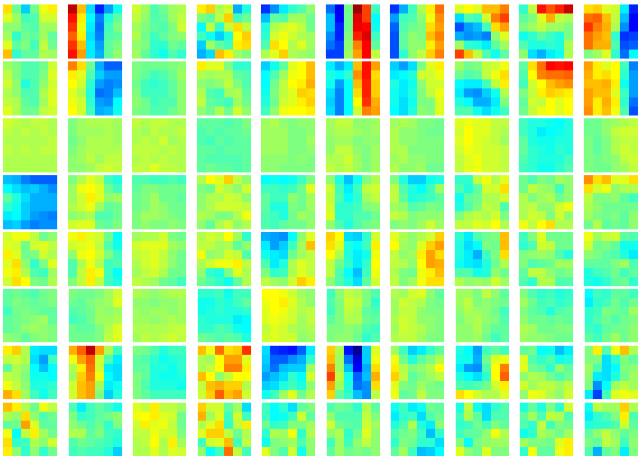
We seek to better understand the reasons for the classification success on the phase imagery. Because CNN B consistently achieved the best classification performance, we investigate its filters in more detail. Fig. 4 shows the learned filters of the three convolutional layers of CNN B. (For a given convolutional layer, each filter uses an identical color scale in which the color green corresponds to zero, warmer colors are positive, and cooler colors are negative.)

TABLE III
CLASSIFICATION PERFORMANCE

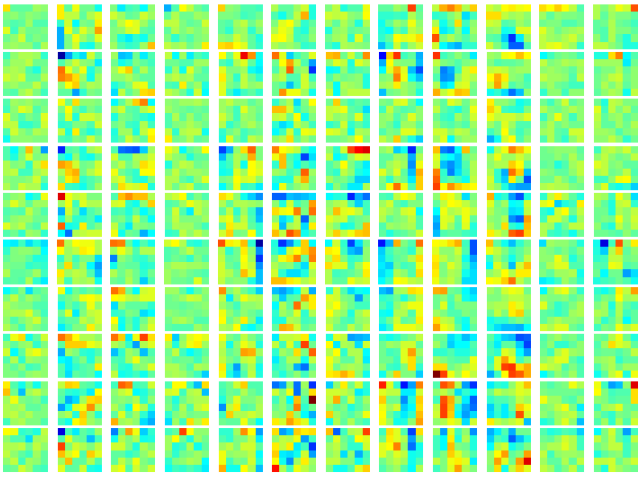
Data Set Code	AUC			
	CNN A	CNN B	CNN C	Ensemble
MAN2	0.857	0.934	0.943	0.946
NSM1	0.613	0.844	0.778	0.773
TJM1	0.807	0.944	0.917	0.936
ONM1	0.835	0.954	0.920	0.947
GAM1	0.862	0.936	0.886	0.928



(a)



(b)



(c)

Fig. 4. For CNN B, the filters of the (a) first convolutional layer, (b) second convolutional layer, and (c) third convolutional layer. (Three-dimensional filters in the latter two layers are grouped columnwise.)

It can be seen that the purpose of the first convolutional layer's filters is ostensibly to locate vertical or horizontal gradients in the input phase imagery. This insight should be contrasted with the finding in [10], which studied CNNs for which the input imagery was SAS *amplitude* images. In that work, it was discovered that the first convolutional layer's filters were effectively acting as bandpass filters (vis-à-vis pixel values) to segment the images into highlight, shadow, and background regions.

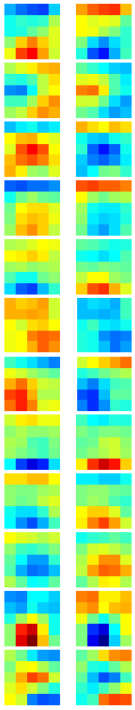
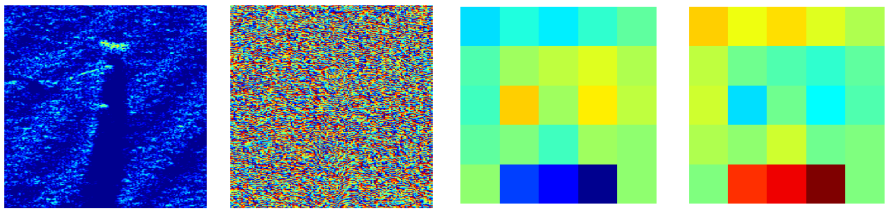


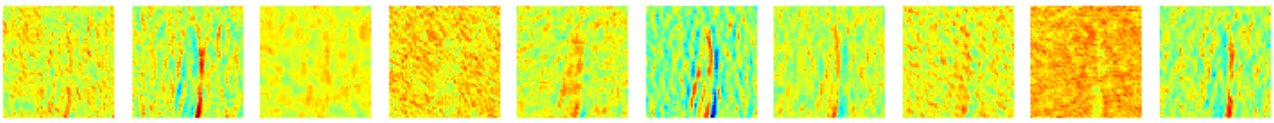
Fig. 5. For CNN B, the weights of the fully-connected layer, displayed such that spatial form is retained (*i.e.,* the "flattening" step has been inverted).

More complicated structural features in the imagery can be isolated when multiple convolutional layers (with nonlinear activation functions) are nested. After the convolutional and pooling layers, the three-dimensional output tensor is "flattened" into a vector to accommodate a fully-connected layer. However, the vector of weights of the fully-connected layer can be reshaped to recover the spatial form destroyed by the flattening. These weights of CNN B are shown in Fig. 5. When presented in this format, one can associate and visualize the relative importance of each spatial region, or "receptive field," of a generic input image. For this CNN, it can be seen that there is not a single dominant region that drives all predictions. Rather, various components will, in general, have the capacity to influence the final class predictions.

Next, we examine some intermediate representations of CNN B for a few specific phase images from the test set. Specifically, we choose to show the intermediate representation of the imagery at the second convolutional layer (prior to evaluation with the sigmoid activation function) because this representation seems to contain the most interpretable structure. In a sense, two convolutional layers are required to transform the phase imagery into a form that humans can readily comprehend. This should be contrasted with a SAS *amplitude* image, which requires no transformations (*i.e.,* convolutional layers) to be understandable.

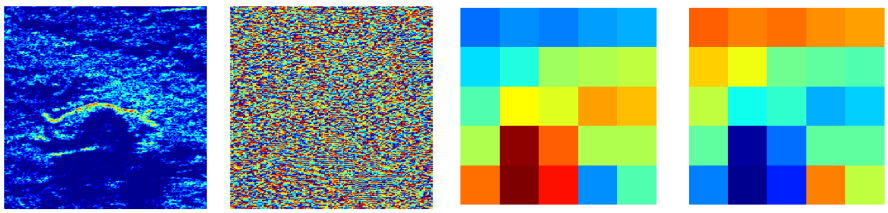


(a) Amplitude image (b) Phase image (c) Contributions to final class predictions

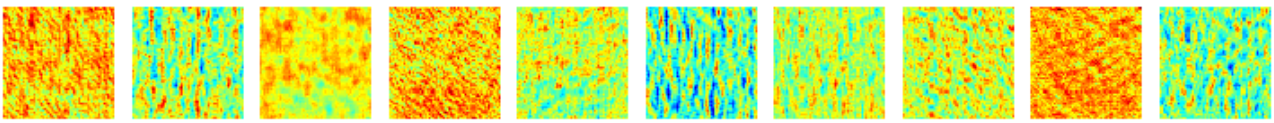


(d) Intermediate representation in CNN B's second convolutional layer

Fig. 6. SAS imagery of a cylinder (target) in a sand ripple field, in the form of its (a) amplitude image and the corresponding (b) phase image. The phase image is the input to the CNN; the amplitude image is shown only for reference. (c) Each receptive field's contribution to the final prediction of belonging to the clutter class (left) and target class (right). The target was classified correctly. (d) The intermediate representation of the phase image after the second convolutional layer, but prior to the activation function, of CNN B.

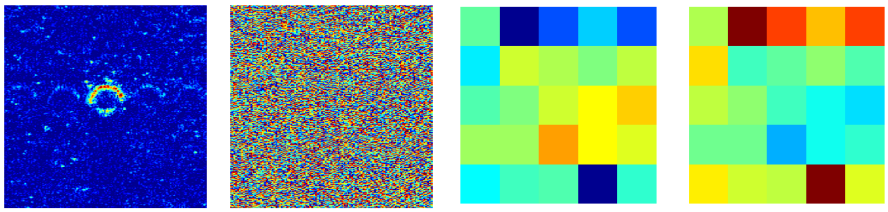


(a) Amplitude image (b) Phase image (c) Contributions to final class predictions

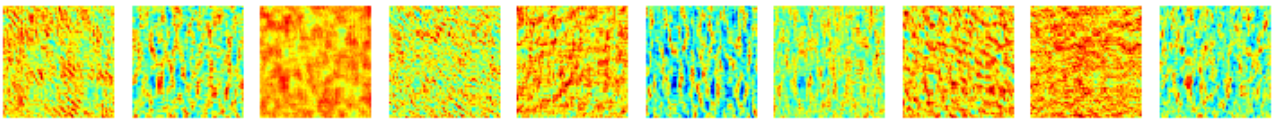


(d) Intermediate representation in CNN B's second convolutional layer

Fig. 7. SAS imagery of a clutter object in the form of its (a) amplitude image and the corresponding (b) phase image. The phase image is the input to the CNN; the amplitude image is shown only for reference. (c) Each receptive field's contribution to the final prediction of belonging to the clutter class (left) and target class (right). The clutter was classified correctly. (d) The intermediate representation of the phase image after the second convolutional layer, but prior to the activation function, of CNN B.



(a) Amplitude image (b) Phase image (c) Contributions to final class predictions



(d) Intermediate representation in CNN B's second convolutional layer

Fig. 8. SAS imagery of a clutter object in the form of its (a) amplitude image and the corresponding (b) phase image. The phase image is the input to the CNN; the amplitude image is shown only for reference. (c) Each receptive field's contribution to the final prediction of belonging to the clutter class (left) and target class (right). The clutter was classified incorrectly. (d) The intermediate representation of the phase image after the second convolutional layer, but prior to the activation function, of CNN B.

We also show the contributions of each spatial location (receptive field) to the final classification predictions. These contributions are the result of inner products between the fully-connected layer's weights and the image responses input to those nodes. That is, these contributions would be summed, added to a bias term, and then passed through a sigmoid function, in order to obtain the final probabilities of belonging to each class. By showing the individual contributions, one can better observe which components of the (phase) image drive the final prediction. (In this series of figures, green corresponds to a value of zero, warmer colors are positively valued, and cooler colors are negatively valued.)

In Fig. 6, we consider a cylindrical target located in a sand ripple field. In the SAS amplitude image, the object's highlight blends in with the background. From the SAS phase image, it is difficult to glean much information. However, by the second convolutional layer, significant structure has been uncovered, as evidenced in Fig. 6(d). For example, some of the filters are effectively delineating the shadow region; the somewhat remarkable thing is that this product has been produced from considering only the *phase* image. It is this sort of feature unearthing that partially explains why the phase can be exploited for target classification. (Results like this also suggest that an additional potential use of phase information can be image segmentation.)

In Figs. 7 and 8, we consider two alarms from the clutter class. In these two cases, the intermediate representations seem to transform the input phase imagery into recognizable textures associated with specific orientations. At earlier layers of the CNN, closer to the original input imagery, the image abstractions are still difficult to understand, while at later layers of the CNN, the meaningfulness of the abstractions again becomes obscured.

In all three figures, it can be observed how different receptive fields influence the predictions to different extents. However, the region in which one would expect to see a shadow (in the amplitude image) – due to the geometry shared by the sensor and proud object – does consistently have a significant impact on the predictions.

Based on these and other preliminary analyses, the information being exploited in the phase imagery for classification appears to arise when the pixel values deviate from a uniform distribution, and more specifically, form non-random spatial structure in the phase. Based on the physics involved, where the phase is tightly coupled to the distance traveled by the sonar signal, this scenario can manifest for different reasons. In [11], spatial correlation in the phase of synthetic aperture *radar* (SAR) imagery was found to be present because of strong reflectors, processing artifacts, and homogeneous surfaces. In [12], evidence of phase structure was seen in SAR images from very strong combined scatterers and their side lobes. In the underwater domain, a discontinuity in the gradient of the phase can be an indication of an abrupt bathymetric change, and the presence of an object proud of the seafloor. Additionally, an acoustically smooth object may produce a structured phase image with a pixel distribution that is not uniform. We hypothesize that another potential source of structure occurs in shadow regions, where the signal levels are so low that deterministic self-noise of the sonar system itself may be visible in the phase.

V. Conclusion

It was demonstrated that the phase information present in complex high-frequency SAS imagery can be exploited for successful object classification. To exploit the information ostensibly hidden in the phase imagery, relatively simple CNNs were trained, "from scratch," on a large database of SAS phase images collected at sea. The filters learned by one of the CNNs were studied, and the intermediate responses from the network for specific input phase images were examined. Hypotheses regarding the sources of information contained in the phase imagery were offered.

Ongoing and future work will seek to better explain the phenomena in the phase imagery driving the classifier predictions. Additional work is being devoted to developing multi-representation CNNs that simultaneously exploit both the amplitude imagery and the phase imagery.

Acknowledgment

This work was supported by the Strategic Environmental Research and Development Program (SERDP).

References

[1] R. Hansen, T. Sæbo, K. Gade, and S. Chapman, "Signal processing for AUV based interferometric synthetic aperture sonar," in *Proc. IEEE OCEANS*, vol. 5, 2003, pp. 2438–2444.

[2] J. Chanussot, A. Hétet, G. Le Merrer, and E. Tireau, "Multispectral decomposition of synthetic aperture sonar images for speckle reduction," in *Proc. Sixth European Conference on Underwater Acoustics (ECUA)*, 2002, pp. 269–274.

[3] S. Silva, S. Cunha, A. Matos, and N. Cruz, "Sub-band processing of synthetic aperture sonar data," in *Proc. IEEE OCEANS*, 2008, pp. 1–8.

[4] T. Marston, J. Kennedy, and P. Marston, "Coherent and semi-coherent processing of limited-aperture circular synthetic aperture (CSAS) data," in *Proc. IEEE OCEANS*, 2011, pp. 1–6.

[5] D. Williams and A. Hunter, "Multi-look processing of high-resolution SAS data for improved target detection performance," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 153–157.

[6] K. Williams, S. Kargl, E. Thorsos, D. Burnett, J. Lopes, M. Zampolli, and P. Marston, "Acoustic scattering from a solid aluminum cylinder in contact with a sand sediment: Measurements, modeling, and interpretation," *The Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3356–3371, 2010.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[8] D. Williams, "The Mondrian detection algorithm for sonar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1091–1102, 2018.

[9] ——, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," in *Proc. 23rd International Conference on Pattern Recognition (ICPR)*, 2016.

[10] ——, "Demystifying deep convolutional neural networks for sonar image classification," in *Proceedings of the Underwater Acoustics Conference*, 2017.

[11] D. Petit, L. Soucille, J. Durou, F. Adragna, H. Oriot, and E. Simonetto, "Spatial phase behavior in SAR images," in *SAR Image Analysis, Modeling, and Techniques IV*, vol. 4543. International Society for Optics and Photonics, 2002, pp. 53–64.

[12] M. Soccorsi and M. Datcu, "Phase characterization of polarimetric SAR images," in *SAR Image Analysis, Modeling, and Techniques IX*, vol. 6746. International Society for Optics and Photonics, 2007.