Application of a Convolutional Neural Network trained with simulated low-frequency Synthetic Aperture Sonar data for classification of buried UXO

Jeroen van de Sande[a], Wyke Huizinga[a], Richard den Hollander[a], Dennis van der Burg[a], Robbert van Vossen[a]

[a]Netherlands Organization for Applied Scientific Research (TNO); Oude Waalsdorperweg 63, 2597 AK, The Hague, The Netherlands; jeroen.vandesande@tno.nl

Abstract: *The application of Artificial Intelligence to target classification problems typically requires substantial amounts of training data. Applications in the underwater domain often suffer from a lack of experimental target data. In addition, operational conditions that affect the target response may be uncertain or unknown. In this paper, a method is described and evaluated that explores the use of Target-In-Environment-Response (TIER) simulations to generate sufficient amounts of training data for the discrimination of proud or (partially) buried Unexploded Ordnance and clutter, using low-frequency Synthetic Aperture Sonar (SAS). As a use case we attempt to discriminate two types of cylindrical dummy targets from clutter. The TIERs of the dummy targets are simulated for a variety of burial depths, orientations and sediment types. Together with field recordings of clutter, they are used for training a Convolutional Neural Network. The trained network is evaluated against discriminating synthetic target data and clutter as well as real target data and clutter, the latter both recorded in field experiments using TNO's MUD system. The use of two SAS data representations has been analysed: conventional time-domain SAS images and Multi-Aspect Acoustic Colour (MAAC) images. The Receiver Operating Curves (ROC) of the CNN for evaluation against an independent set of synthetic data shows an average Area-Under-Curve of 0.99 for both data types. Application on real data shows a performance reduction for SAS images to an AUC of 0.86, which is considered reasonable. For MAAC images, however, the performance breaks down to an AUC of 0.64. A mismatch between target model and actual target, in conjunction with CNN overfitting, is suspected to be the main reason for this performance drop. Suggestions are made to extend the training set with simulations that include several variations in the target models, in order to reduce the sensitivity to the precise target properties.*

Keywords: *Synthetic Aperture Sonar, low-frequent, UXO, target classification, Target-in-environment-Response, data simulation, Deep Learning, Convolutional Neural Network*

1. INTRODUCTION

1.1. Problem

The application of deep learning typically requires substantial amounts of training data with sufficient variation in order to train good classifiers and provide reliable target recognition in operational environments. Particularly for military operations, several challenges hinder the successful application of deep learning:

- A lack of availability of high volumes of relevant target data for training, leading to under specification [1].
- Biases or data shifts in training sets caused by environmental conditions, which may have a significant impact on target signature and background reverberation
- Lack of variability in deployment characteristics, such as target orientations and burial depths
- Unavailability of particular (enemy) targets in experimental data

As a consequence of these challenges, it is not guaranteed that a trained target recognition approach is applicable to the prevalent operational conditions. To overcome these challenges, an attempt is made to use large amounts of synthetic target data for training a classifier.

1.2. Use case

The Mine Underground Detection (MUD) system [2] has been developed to detect and localize buried objects in the seabed in inshore environments, such as harbours. It uses an interferometric low-frequency Synthetic Aperture Sonar (LF-SAS) as primary sensor to detect proud and buried objects, which operates at frequencies below 30 kHz. During several field experiments, recordings have been made of mine-like targets. Among those are two dummy UXO: NL-REF and NL-CYL. NL-REF is a 0.5 m x 1 m aluminium cylinder with internal structure and filled with water. NL-CYL is a 0.3 x 0.6m solid aluminium cylinder. It is a replica of the target used by [3]. See Fig. 3.

In the use case described in this paper synthetic target data of the two dummy UXO are generated with a developed target simulation framework. They are used in conjuction with real clutter images to train a Convolutional Neural Network (CNN). The trained network is evaluated against real target images and real clutter images. The reliability of the classifier is considered through self-assessment. .

2. GENERATION OF SYNTHETIC TARGET-IN-ENVIRONMENT DATA

Synthetic Target-in-Environment images are generated according to the flowchart given in Fig. 1. TIER simulation comprises the computation of a generic Look-up Table (LUT) containing aspect and frequency dependent target responses, including the interaction between the target and a sediment interface layer. With this table, large numbers of measurement realisations can be generated for the particular target in which

the characteristics of the used sonar are applied and in which target deployment locations and orientations can be varied. Both blocks will be described in more detail in this section.

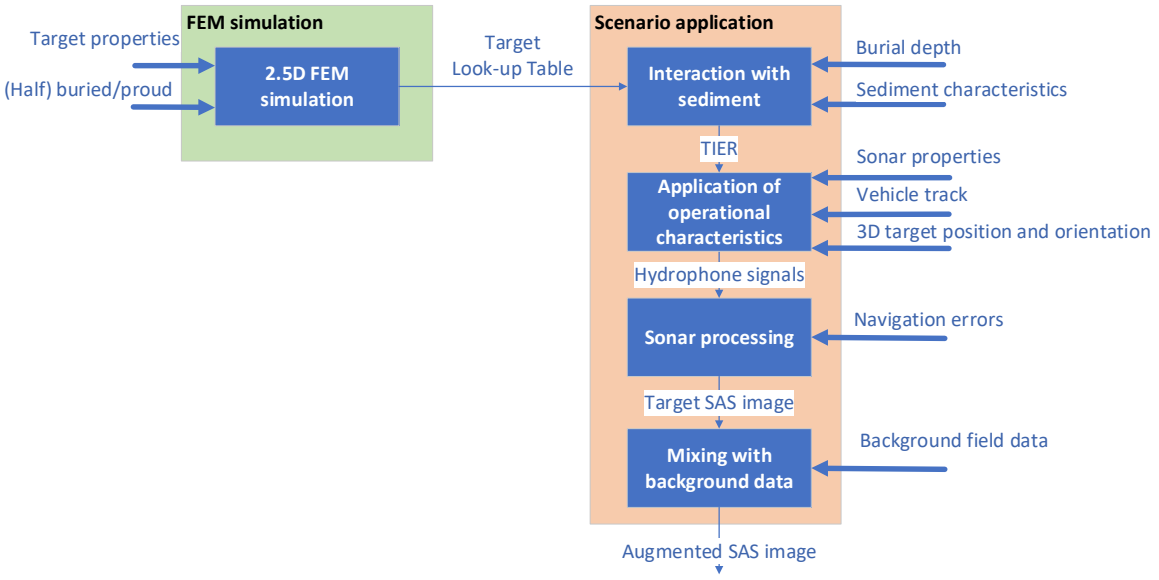

Fig.1: Flowchart of the simulation framework for generating augmented TIER images

2.1. Target-in-environment response simulation

In FEM simulations the aspect and frequency dependent target response is computed, including the interaction between the target and the sediment interface layer(s). TNO-AXISCAT [4] is a high-fidelity simulation capability that is used for this purpose (Fig. 2). It is based on finite-element modelling (FEM) and simulates the full elastic response of axisymmetric objects, such as cylinders.

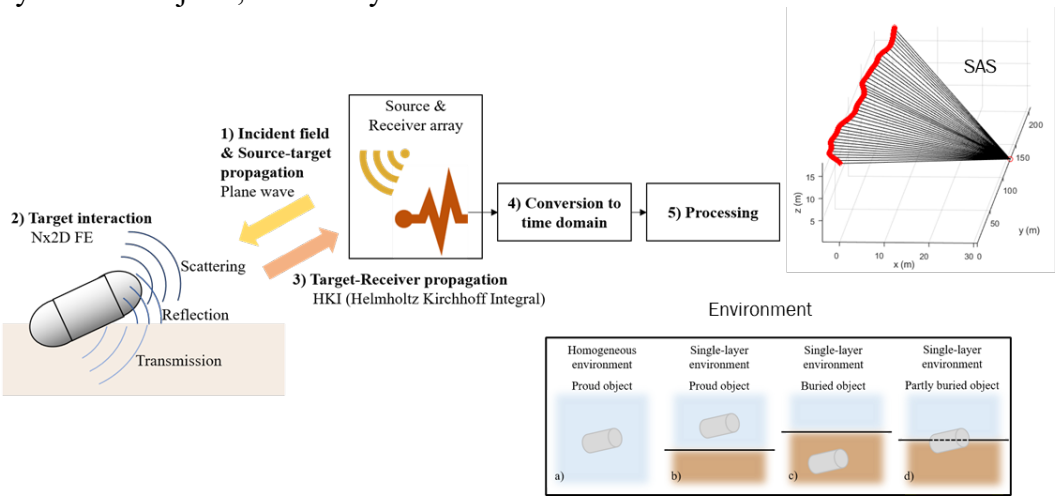

Fig.2: Overview of TNO-AXISCAT for simulating the Target-in-Environment Response

2.2. Scenario application

The generated LUT is used in a computationally cost-efficient way to generate large numbers of image realisations of the same target under different operational and

deployment conditions. This provides the ability to create a significant training data set with randomized variations in responses, which are indicated in Fig. 1.

For the generation of a single SAS image, a vehicle trajectory is simulated along a target at a given location, orientation and burial depth. For each ping, the target response is extracted from the LUT. The responses are corrected for propagation loss and for the angle and frequency dependent sonar response. The responses are converted to time domain for all receiving hydrophones, resulting in an entire sonar recording for the given vehicle track. The resulting data is processed with the same sonar processing chain that is used for processing field data. Potential navigation errors can be added to avoid too perfect target image reconstruction. Ultimately, the target's SAS image is coherently blended with random background SAS images extracted from field recordings of several environments. Fig. 3 shows examples of real and synthetic SAS and Multi-Aspect Acoustic Colour (MAAC) images for two realisations of the example targets.

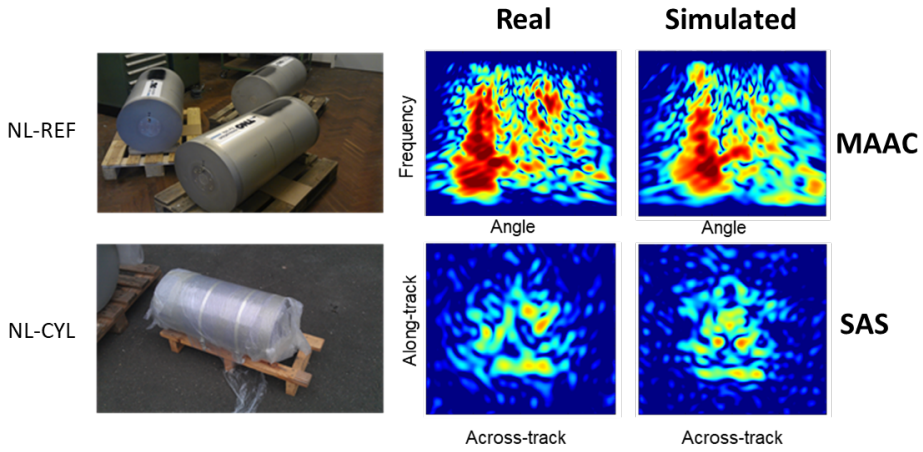

Fig.3: Real (top center) and simulated (top right) MAAC image of NL-REF target (top left). Real (bottom center) and simulated (bottom right) SAS image of NL-CYL target (bottom left).

3. TARGET CLASSIFICATION APPROACH

For the classification between target and clutter, a CNN is used. The target class comprises the collection of NL-REF and NL-CYL images. For training, only the synthetic images are used. The clutter class consists of clutter images recorded in field experiments. Two CNNs are trained and evaluated, each using a different data representation. One CNN is trained on conventional time-domain Synthetic Aperture Sonar (SAS) images, the other CNN on Multi-Aspect Acoustic Color (MAAC) images. The results of both CNNs are assessed individually.

3.1. Basic CNN architecture

The architecture of the used CNN is shown in Fig. 4. The input size of the images are scaled to [n×m]=256×256. The dense layer of the single input architecture has size [k]=576. The models are trained with an Adam optimizer [5], using a learning rate scheduler where the learning rate is reduced by a factor $\gamma<1$ with each epoch. In our experiments we used $\gamma=0.95$. An initial learning rate of 0.001 is used; the model is trained for 100 epochs with a batch size of 60 on a single GPU.

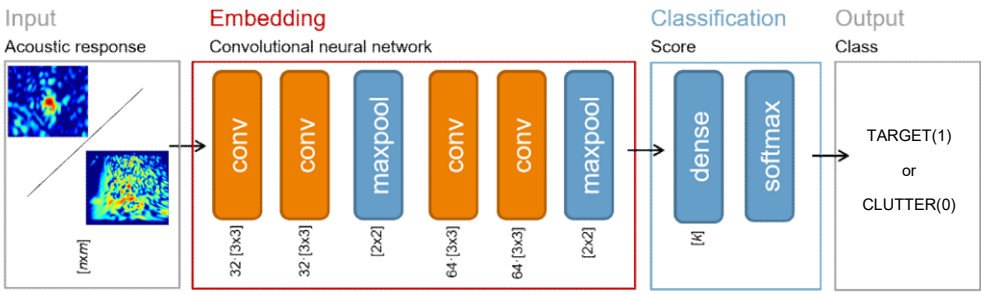

Fig.4: Basic architecture of the CNN used for classification of target and clutter.

3.2. Classifier reliability

Deep neural networks have the tendency to be overconfident in classifying test samples that are outside the distribution of the training set, producing confidence values as high as those for samples that are within the distribution [6][7]. Since the confidence (or uncertainty) of a classifier becomes important in safety critical applications, the investigation of neural network uncertainty is of high relevance. There exist methods to improve the uncertainty estimates in classifier output, like Bayesian neural networks, where the epistemic (model) uncertainty is estimated by Markov Chain Monte Carlo methods or variational inference [8]. A simpler and computationally less demanding method is Monte Carlo dropout [9].

In this paper another conceptually simple method is used that constructs ensembles of multiple networks [10]. In this ensemble approach, the network is trained several times using different random weight initializations. The trained networks will reflect uncertainty by their spread in classification outputs. It has been observed in imaging experiments that ensembles can produce better uncertainty estimates than Bayesian neural networks or Monte Carlo dropout [11][12].

4. EXPERIMENTS AND RESULTS

Training data consisted of ~1000 synthetic NL-CYL and NL-REF target views and ~500 real clutter images. The trained classifier was evaluated on an independent set of synthetic data and real clutter. This was the baseline result. Additionally the classifier was evaluated on the set of real targets and real clutter. The set of real targets consisted of a total of 78 NL-CYL and NL-REF images with different target orientations, grazing angles and burial depths. The set of real clutter was an independent set of 153 images.

4.1. Application to synthetic target data

Fig. 5 shows ROC curves of the baseline test:

- Training: synthetic targets vs. real clutter
- Evaluation: synthetic targets vs. real clutter (independent sets)

The left image is the result of the CNN classifier for SAS images, the right image is the result of the CNN for MAAC images. The Area Under Curve (AUC) is around 0.99 for both cases, which is a good result.

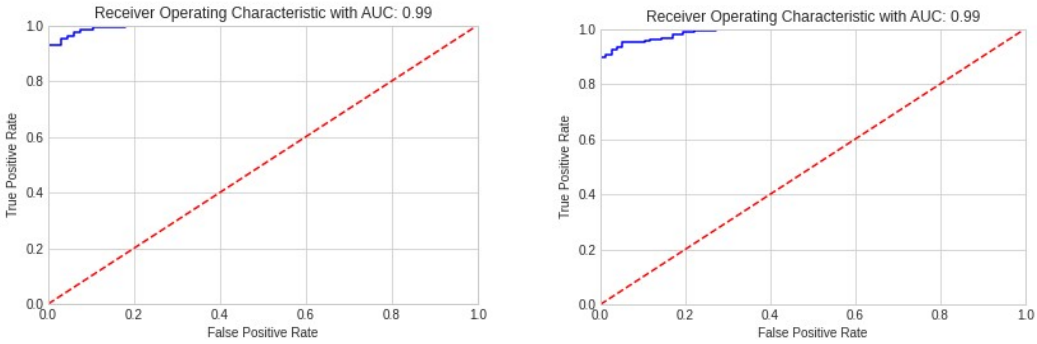

Fig.5: ROC curves (blue) of classifier evaluation on synthetic target data vs. real clutter using SAS images (left) and MAAC images (right). For comparison, the result of a random classifier (red) that cannot distinguish between targets and clutter is shown.

4.2. Application to real target data

Fig. 6 shows ROC curves of the real target test:

- Training: synthetic targets vs. real clutter
- Evaluation: real targets vs. real clutter (independent set)

The left image is the result of the CNN classifier for SAS images, the right of the CNN for MAAC images. The result of the SAS classifier with an AUC of 0.86 is considered to be acceptable for the given test scenario. The result of the MAAC classifier with an AUC of 0.64 is considered to be poor.

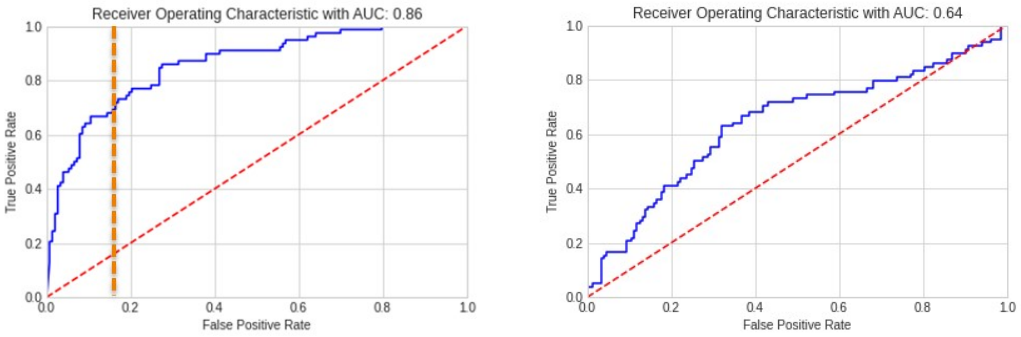

Fig.6: ROC curves (blue) of classifier evaluation on real target data vs. real clutter data using SAS images (left) and MAAC images (right).

To provide insight in the difference in classification performance between NL-CYL and NL-REF, a sample point is taken at the dashed orange line in the left ROC-curve of Fig. 6. At this point the False Alarm Rate (FAR) equals 0.17, indicating that at this confidence threshold 17% of the clutter contacts are falsely classified as targets. At this FAR, 100% of the NL-CYL targets are classified correctly and 70% of the NL-REF targets are classified correctly.

4.3. Classifier reliability

Fig. 7 shows the self-assessment results of a single classifier and an ensemble classifier for SAS images, respectively. These plots have been generated from the experiment where SAS data of the real targets was used in evaluation (Fig. 6 left). The calibration curves should ideally follow the dashed, orange line. For example, of all the samples that received a classifier predicted probability of 0.8 of being a target, the fraction of target samples in this subset should also be 0.8 for a well calibrated classifier, indicating that the predicted probability represents the actual probability of seeing a target.

The use of an ensemble classifier instead of a single classifier overall shows an improvement in reliability. The fact that the calibration curves in general still differs from the optimum line may partly be assigned to a remaining classifier calibration error. However, the small amount of real target data in evaluation potentially also plays a role.

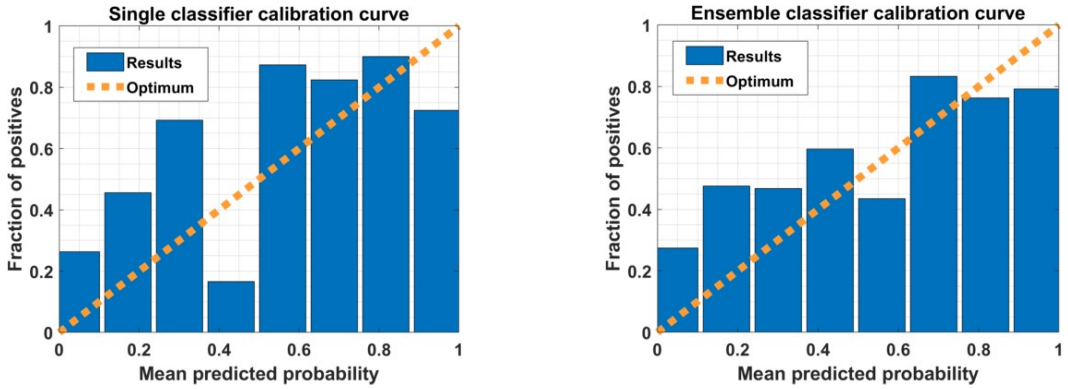

Fig.7: Calibration curve of the single classifier (left) and the ensemble classifier (right) from the classifier using SAS images of the real target data.

5. CONCLUSIONS

The use of synthetic data for training a deep neural network for the classification of mine-like targets has shown some preliminary potential. The results of the classification of real targets and real clutter using SAS images are reasonable. The classification results using MAAC images are still poor, however. The self-assessment performance of the classifier is of a reasonable level. The benefit of using an ensemble classifier as opposed to a single classifier is observable.

The hypothesis for the difference in performance between MAAC and SAS images is the fact that MAAC images contain more detailed information on target scattering than SAS images (without phase). Deviations in the synthetic model of the target may therefore be more prone to overfitting. Improving the target model to create a better match with the actual target would potentially improve results. However, fitting a classifier to this exact target, will not result in an operationally applicable solution; it may be required to force the classifier to generalize in order to derive a more generic picture of a cylindrical target. Consequently, some recommendations are to reduce the amount of detail in the images, e.g. by down sampling the images, or to extend the training set with additional simulations where the target properties are modified.

6. ACKNOWLEDGEMENTS

This work has been carried out as part of a research programme on reliable target recognition, funded by TNO's Appl.AI program and the NL MoD and supported by the Defence Material Organisation (DMO) and the Royal Netherlands Navy Command (RNLN).

REFERENCES

[1] **D'Amour, A., Heller, K., Moldovan, et al.,** Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research,* vol. 23, 2020.

[2] **van Vossen, R., van de Sande, J., Duijster, A., van der Burg, D., Mulders, I., & Beckers, G.**, Seabed characterization with multi-band interferometric sonar, Proceedings of Meetings on Acoustics ICUA, vol. 40, p. 070009, 2020.

[3] **Williams, K. L., Kargl, S. G., Thorsos, E. I., Burnett, D. S., Lopes, J. L., Zampolli, M., & Marston, P. L.**, 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, pp. 3356–3371, 2010.

[4] **Zampolli, M., Tesei, A., Jensen, F. B., Malm, N., & Blottman III, J. B.,** A computationally efficient finite element model with perfectly matched layers applied to scattering from axially symmetric objects. *The Journal of the Acoustical Society of America,* vol. 122, pp. 1472–1485, 2007.

[5] **Kingma, D. P., & Ba, J., Adam**, A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.

[6] **Hein, M., Andriushchenko, M., & Bitterwolf, J.,** Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 41–50, 2019.

[7] **Gerg, I. D., & Monga, V.**, Preliminary Results on Distribution Shift Performance of Deep Networks for Synthetic Aperture Sonar Classification, OCEANS 2022, Hampton Roads, VA, USA, pp. 1-9, 2022.

[8] **Jospin, L.V., Laga, H., Boussaid, F., Buntine, W. and Bennamoun, M.**, Hands-On Bayesian Neural Networks—A Tutorial for Deep Learning Users, IEEE Computational Intelligence Magazine, vol. 17, no. 2, pp. 29-48, 2022.

[9] **Gal, Y., & Ghahramani, Z.**, Bayesian convolutional neural networks with Bernoulli approximate variational inference. arXiv preprint arXiv:1506.02158, 2015.

[10] **Lakshminarayanan, B., Pritzel, A., & Blundell, C.**, Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems, vol. 30, 2017.

[11] **Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Snoek, J.,** Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems,* vol. 32, 2019.

[12] **Gustafsson, F. K., Danelljan, M., & Schon, T. B.**, Evaluating scalable bayesian deep learning methods for robust computer vision methods for robust computer vision. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Seattle, WA, USA, pp. 1289-1298, 2020.