Hand-Crafted Feature Based Classification against Convolutional Neural Networks for False Alarm Reduction on Active Diver Detection Sonar Data

Matthias Buß
University of Wuppertal
Wuppertal, Germany
matthias.buss@uni-wuppertal.de

Yannik Steiniger
ATLAS ELEKTRONIK GmbH
Bremen, Germany
yannik.steiniger@atlas-elektronik.com

Stephan Benen
ATLAS ELEKTRONIK GmbH
Bremen, Germany
stephan.benen@atlas-elektronik.com

Dieter Kraus
City University of Applied Sciences
Bremen, Germany
dieter.kraus@hs-bremen.de

Anton Kummert
University of Wuppertal
Wuppertal, Germany
kummert@uni-wuppertal.de

Dietmar Stiller
*WTD 71**
Eckernförde, Germany
dietmarstiller@bundeswehr.org

***Abstract*—This paper provides a method for false alarm reduction on active diver detection sonar data by means of contact classification. Two different classification techniques are compared, on the one hand a classical hand-crafted feature based method and on the other hand a deep learning approach which inherently includes an automated feature extraction. For the classical method 53 hand-crafted features are extracted from the sonar contacts and used as inputs for a feedforward neural network (FNN). The inputs for the convolutional neural networks (CNNs) are two-dimensional sonar images (level vs. bearing and time/range) of the contacts. The performance of the classifiers is compared to that of the standard active signal processing by means of Receiver-Operating-Characteristic (ROC) curves. It is shown that both classification techniques outperform the standard active signal processing in which the detector threshold represents the only adjustable parameter.**

***Index Terms*—Active Sonar, False Alarm Reduction, Contact Classification, Neural Networks, Deep Learning**

I. Introduction

In the last decades the requirements for active sonar applications changed essentially. While in the past the detection and classification of targets was done manually by sonar operators nowadays the systems should work more and more automatically. Ideally, a modern sonar system should reliably detect, track and classify threats and report an alarm. Therefore, the biggest challenge is to achieve a high probability of detection and simultaneously a low false alarm rate. In common standard high frequency active sonar applications generally two different pulse types are used; on the one hand broadband frequency modulated (FM) pulses and on the other hand narrow-band continuous wave (CW) pulses. In case of linear or hyperbolic frequency modulated (LFM/HFM) pulses usually only the signal-to-noise ratio (SNR) of the contacts is used as measure of reliability whereas for CW pulses in addition to the SNR also the Doppler information is considered. However, it is known that echoes contain more information that can be used to assess their relevance and hence improve the detection performance [1]–[5]. In previous work [6] it is shown that the extraction of features of the contacts in combination with supervised machine learning algorithms is suited to reduce the false alarm rate of an active sonar system. For this purpose, a lot of hand-crafted feature engineering is required. However, the high computational power of today's available computers pushed the deep convolutional neural networks (CNNs) in the foreground of many classification tasks [7]–[11]. These networks do not require manual feature engineering since they automatically extract features out of labeled input signals or images.

In this work the performance of a hand-crafted feature based feedforward neural network (FNN) is compared with that of convolutional neural networks regarding their suitability for reducing the false alarm rate without degradation of the probability of detection. The algorithms are applied to recorded data of active diver detection trials that were carried out in cooperation between the WTD 71 and ATLAS ELEKTRONIK. The trials were conducted with the "Cerberus" diver detection sonar developed by ATLAS ELEKTRONIK UK. It should be noted that all results are based on the transmission of FM-pulses which are processed with an experimental signal processing in MATLAB. In total 53 hand-crafted features from different categories are extracted from the contacts and represent the inputs of the FNN. The FNN consists of one hidden layer and an output layer with two neurons for binary classification (target contact or false alarm). For the use of CNNs a region of interest (ROI) of the two-dimensional sonar images (level vs. bearing and time/range) is extracted for each contact. Two different CNNs are considered; a shallow CNN trained from scratch and the pre-trained deep network VGG-16 [12]. In this way the influence of the depth of the networks

*Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71).

978-1-5386-4814-8/18/$31.00 ©2018 IEEE

is investigated. The classification performance is assessed by means of Receiver-Operating-Characteristic (ROC) curves and the generalization respectively the robustness of the classifiers is proved by testing the algorithms with unseen data recorded in different environments.

The paper is structured as follows. In Section II the modified active signal processing chain for contact classification is presented. A method for labeling the contacts and information about the considered datasets are described in Section III. The different classification algorithms and the corresponding inputs are introduced in Section IV followed by a discussion of the achieved results in Section V. Finally, a summary and tasks for future work are provided in Section VI.

II. Modification of Active Signal Processing

For contact classification the standard active signal processing chain has to be extended by a feature extraction and classification step. The modified signal processing chain is shown in Fig. 1. The gray boxes illustrate the signal processing steps of a standard active sonar. At first the raw data is applied to a beamforming algorithm in which the azimuthal direction of incidence of the received signals is determined. In a second step a matched filter that maximizes the SNR is applied to the beamformed data. For subsequent detection with e.g. a threshold detector the data have to be normalized. The normalization is divided into two parts; the estimation of the background noise and the normalization to that. This results in an estimate of the SNR which is used for detection. After applying a threshold detector neighboring threshold crossings are merged to contacts. The detector threshold represents the only adaptable parameter in the standard signal processing which affects the detection performance. Finally, the tracking algorithm associates the contacts from successive pings to tracks. The contacts as well as the tracks are displayed in the detection display.

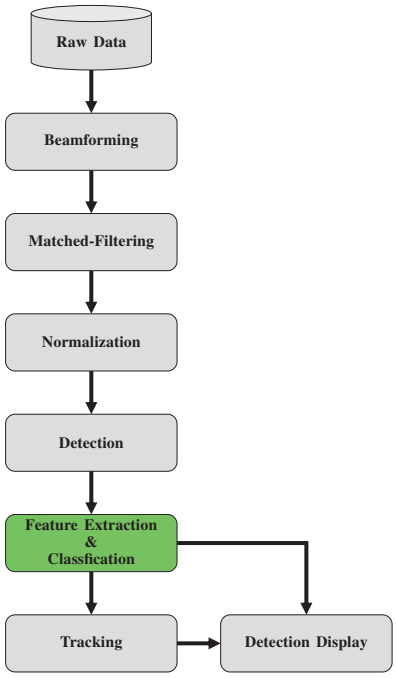


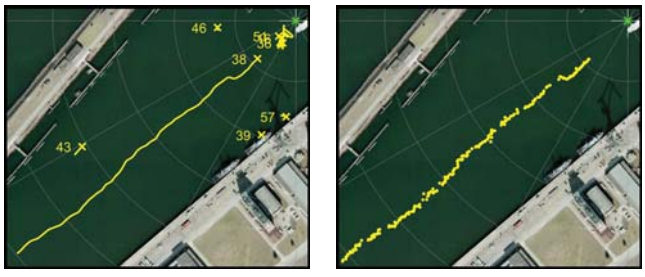
Fig. 1. Modified active signal processing chain for false alarm reduction.

The modification of the standard signal processing is emphasized by the green block. For the use of classical handcrafted feature based classification algorithms it is necessary to define and extract features. However, in case of convolutional neural networks no features have to be defined since the calculation of discriminative features is part of the training process. Both classification approaches provide a probability for class affiliation which in combination with a threshold could directly be used for reducing the number of contacts. Alternatively, the classification results can be used within the following tracking algorithm as measure of reliability.

III. Contact Labeling and Data

For the training of supervised machine learning algorithms labeled input data are required. In active sonar applications a ground truth of the targets trajectory is often not absolutely available so that the true target contacts have to be found manually. However, this process is very time-consuming and can, due to a large amount of false contacts, lead to errors. Hence, in this work an almost automated method is applied. The track that belongs to the diver is used as ground truth and all corresponding contacts that are chosen by the tracker are labeled as "Diver Contact". All remaining contacts are labeled as "False Alarm". Nevertheless, this method does not guarantee that all labeled diver contacts are actually caused by the diver but if no exact ground truth is available this errors can also occur when the contacts would be labeled manually. Fig. 2a illustrates the tracking results at the end of a diver detection trial in which a diver was approaching the Cerberus sonar whose position is marked by the green star. The positions of all contacts that belong to the track of the approaching diver (No. 38) are displayed by the yellow points in Fig. 2b.

In this work three runs from different diver detection trials are considered. For all datasets the pulse parameters, the number of considered pings as well as the number of diver contacts and false alarms are given in Table I. The datasets D_1H_1 and D_2H_1 are part of the same test series, which means



(a) (b)

Fig. 2. Tracking results of a diver detection trial (a) and corresponding contacts that belong to the track (No. 38) of the approaching diver (b).

TABLE I
INFORMATION FOR CONSIDERED DATASETS

Shortcut	Pulse Parameters				# Pings	# Diver Contacts	# False Alarms
	Type	*Center Frequency*	*Bandwidth*	*Length*			
D_1H_1	HFM	100 kHz	12 kHz	50 ms	346	240	28896
D_2H_1	HFM	100 kHz	12 kHz	50 ms	442	328	52282
D_3H_2	LFM	100 kHz	20 kHz	100 ms	179	138	33010

that both recordings took place at the same day in the same harbor (H_1). Furthermore, in both runs the pulse parameters and scenarios were the same; a diver moves away from the sonar, turns and approaches the sonar. In trial D_2H_1 the speed of the diver was a little bit lower than in D_1H_1 which results in more transmissions and hence more contacts. Dataset D_3H_2 differs from the previously described datasets in several ways. On the one hand the trial took place in another harbor (H_2) and on the other hand different pulses were transmitted. In this trial a diver was approaching the Cerberus sonar so that the scenario is comparable to the last part of the other two datasets.

Especially for the training of CNNs in general a large amount of labeled input images per class is required to learn discriminative features. From Table I it can be seen that many false alarms but only a few diver contacts are available in all three datasets. One method to overcome this problem is an artificial augmentation of the data. However, an augmentation of the data is not considered in this work but should be analyzed in the future.

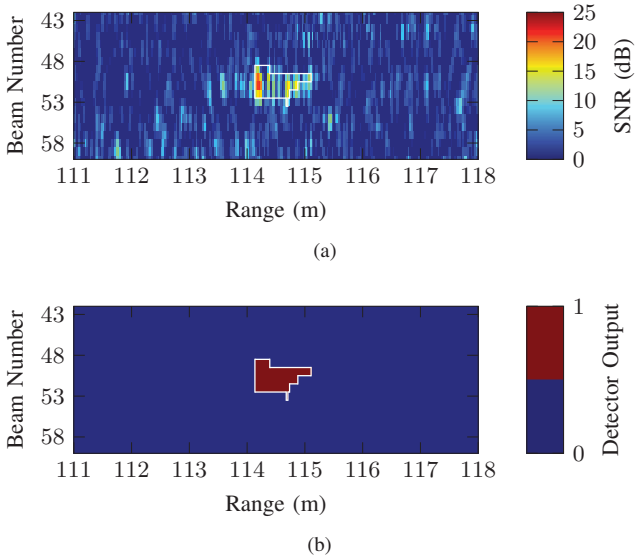


Fig. 3. Snapshots of a diver contact after the two signal processing steps "Normalization" (a) and "Detection" (b).

IV. Classification Algorithms and Inputs

A. Feedforward Neural Network

The hand-crafted feature based classification is considered by means of a FNN. As mentioned in Section II the standard active signal processing chain is extended by a feature extraction for subsequent classification. Some examples for the extracted features can be given with respect to the snapshot of the two-dimensional data of a diver contact presented in Fig. 3. On the horizontal axis a section of 7 m in range direction is shown. In the vertical domain the signals of 19 adjacent beams are presented. Fig. 3b represents the output of the detector. From this data the maximum extent in range direction as well as in beam direction are extracted and represent two features. In this example the extent in range direction is $\approx$ 1 m and the extent in beam direction is 5 Beams. Fig. 3a illustrates the corresponding normalized data, in which the contact pixel of the diver contact are highlighted by the white edges. From the highlighted area e.g. the maximum and the mean of the SNR values are extracted as another two features. In the standard active signal processing only the SNR is used for detection so that a contact only occurs if the SNR is above the detector threshold. Hence, the maximum SNR of a contact represents the reference feature of the standard signal processing. In the following the maximum SNR of a contact is referred to as "Contact SNR".

In total 53 features are defined based on the level, extent, spectrum, sub-bands, and statistics of the contact signals. Since a detailed description of each individual feature would go beyond the scope of this work it should be noted that a variety of the extracted features can be found in [1]–[5]. For the i-th contact the feature extraction results in the feature vector

$$\mathbf{x}_i \in \mathbb{R}^{1\times 53} \tag{1}$$

and for a whole dataset in the feature space

$$\mathbf{X} = (x_{ij}) \quad \begin{matrix} i = 1, 2, \ldots, N_C \\ j = 1, 2, \ldots, 53, \end{matrix} \tag{2}$$

where i and j index the contacts and features respectively. Furthermore, for each contact the class label is represented in the class vector $\mathbf{t}$ with the elements

$$t_i \in \{c_1, c_2\}, \tag{3}$$

where c_1 and c_2 are equivalent to the classes "Diver Contact" and "False Alarm". The feature space as well as the class vector are used to train the FNN. According to the universal approximation theorem, a FNN consisting of one hidden layer and a sufficient number of neurons is able to approximate any continuous function with sufficient accuracy under the constraint that the activation function is limited, continuous and nonconstant [13]. Regarding this, the FNN in this work

consists of one hidden layer and the activation function is the hyperbolic tangent which fulfills the aforementioned constraints. The structure of the used FNN is illustrated in Fig. 4. It can be seen that the input of the network consists of 53 input variables which represent the extracted features. These are forwarded to 19 neurons in the hidden layer. This number is the result of a study in which the number of neurons was varied from one to 100 in steps of one for finding the best performing setup. The output layer of the network consists of two neurons and uses the softmax function for the calculation of a probability of class affiliation. For a new contact represented by its feature vector $\mathbf{x}_n$ the probability for belonging to class c_1 and c_2 is given by

$$y_1(\mathbf{x}_n) = p(c_1|\mathbf{x}_n) \tag{4}$$

respectively

$$y_2(\mathbf{x}_n) = p(c_2|\mathbf{x}_n) \tag{5}$$

and the sum of both probabilities is one so that

$$y_2 = 1 - y_1. \tag{6}$$

Since the performance of the classification algorithm is measured by means of an ROC curve a threshold $\theta \in [0, 1]$ is introduced to adjust the final classification result with

$$\hat{t}_n(\mathbf{x}_n, \theta) = \begin{cases} c_1 & y_1(\mathbf{x}_n) \geq \theta \\ c_2 & \text{otherwise.} \end{cases} \tag{7}$$

The FNN in this work is realized with the Neural Network Toolbox in MATLAB. For training the network the feature space is normalized to get feature values in the interval $[-1, 1]$ by the MATLAB function `mapminmax()` with

$$\tilde{x}_{ij} = 2\frac{x_{ij} - \min(\mathbf{v}_j)}{\max(\mathbf{v}_j) - \min(\mathbf{v}_j)} - 1, \tag{8}$$

where $\mathbf{v}_j$ contains the N_C feature values of the j-th column in the feature space of the training dataset. Furthermore, the scaled conjugated gradient algorithm with the MATLAB default settings is used and early stopping is applied when the validation error does not drop in six consecutive epochs.

B. Convolutional Neural Networks

In addition to the hand-crafted feature based FNN two different CNNs are considered. For each contact a region of interest (ROI) of the normalized two-dimensional sonar images is extracted. An example for the extraction of the input data can be given by the diver contact shown in Fig. 3a. The normalized data are stored for a section of ± 2 m and ± 5 Beams around the pixel of the weighted contact centroid. This results in an input image with a size of 142×11 pixel which is shown in Fig. 5. It can be seen that the contact and its surrounding area build the ROI that represents the input of the CNNs. The intensities of the input images are the linear SNR values of the normalized data.

In this work on the one hand a shallow CNN which is trained from scratch and on the other hand the pre-trained deep

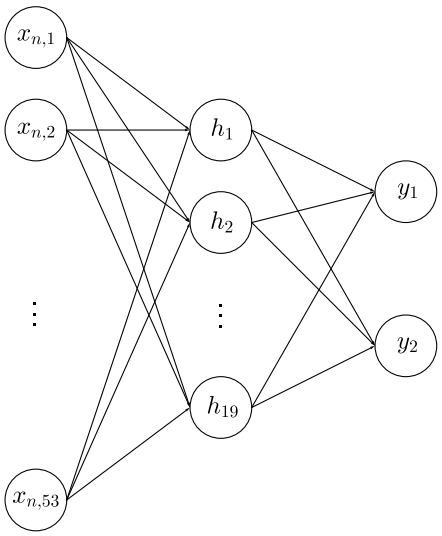


Fig. 4. Structure of trained FNN consisting of 53 inputs, a hidden layer with 19 neurons and an output layer with two neurons for binary classification.

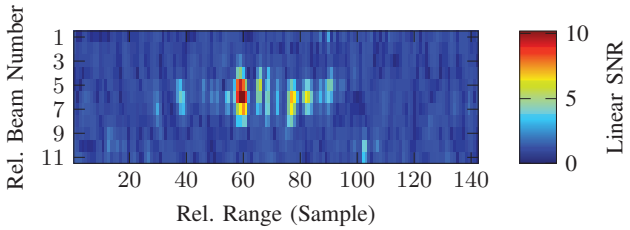


Fig. 5. Example for an input image for the use of CNNs.

VGG-16 network are considered. The structure of the shallow CNN is illustrated in Fig. 6a. It can be seen that the network consists of one convolutional layer, one average pooling layer and finally a FNN with one hidden layer and one output layer for binary classification. The dimensions of each layer are the result of a study in which the kernel size, the number of kernels and the number of neurons in the FNN was varied to find the setup which performs best for this type of data. The activation function is the Rectified Linear Unit (ReLU) and the network is trained in MATLAB with the training options

- `sgdm` "Stochastic Gradient Descent with Momentum"
- `MaxEpochs` : `60`
- `MiniBatchSize` : `1500`
- `InitialLearnRate` : `0.4`
- `LearnRateSchedule` : `piecewise`
- `LearnRateDropPeriod` : `20`
- `LearnRateDropFactor` : `6.`

Since the output layer of the network consists of a softmax function, the result of the classification with the CNN is comparable to that of the FNN described in (4)-(7).

The structure of the VGG-16 network is illustrated in Fig. 6b. It can be seen that the network is much deeper than the previous one. If a convolutional layer with the same

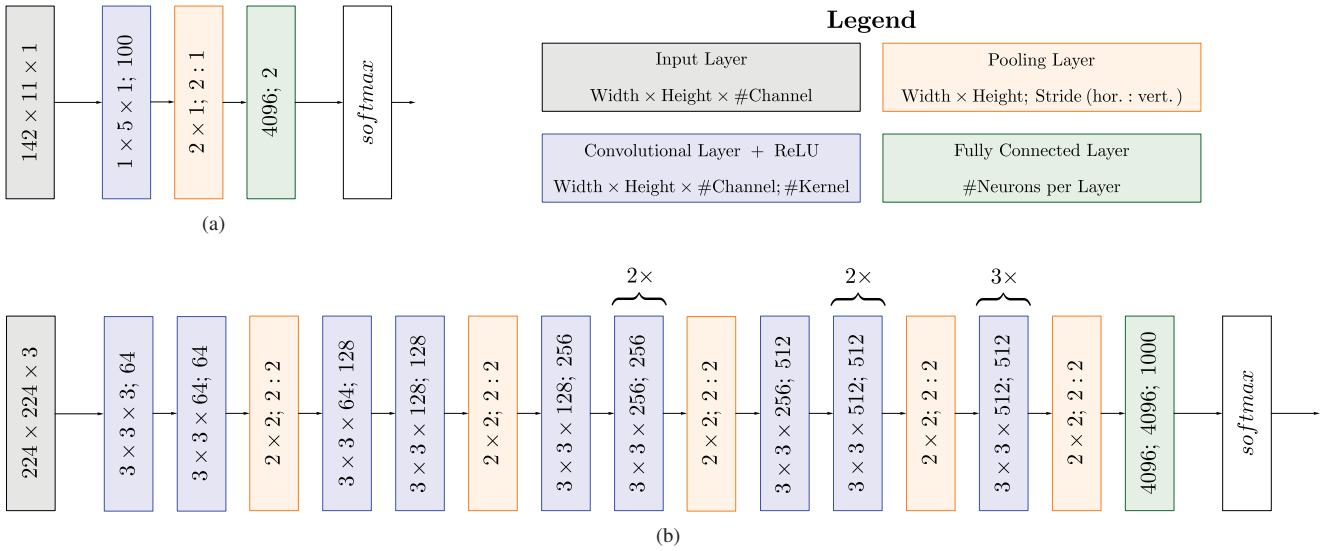


Fig. 6. Structure of shallow CNN trained from scratch (a) and pre-trained VGG-16 (b). The number above a bracket indicates how often a step is repeated.

configuration is executed several times consecutively this is indicated by the number of repetitions above the curly bracket to keep a clear overview. It should be mentioned that the fully connected layers of the VGG-16 apply dropout with a factor of 0.5 to avoid overfitting. For the use of this network the input images have to be resampled to a size of 224×224. Furthermore, the resampled indexed images have to be converted into R-G-B channels which is realized with the MATLAB function `imwrite()` by using the jet colormap. Since the VGG-16 network was initially trained to distinguish between 1000 classes, the output of the fully connected layer has to be redefined for binary classification. The training options for the transfer learning are set to

- `sgdm` "Stochastic Gradient Descent with Momentum"
- `MaxEpochs` : 3
- `MiniBatchSize` : 100
- `InitialLearnRate` : 0.0004.

For the new defined output layer the options

- `WeightLearnRateFactor` : 20
- `BiasLearnRateFactor` : 20

ensure that the learning rate for the training is not too small. Therefore, the value is chosen to be 20 times higher than the learning rate for the previous layers. A comparison of the options for the training of the shallow CNN and the VGG-16 reveals that the initial learning rate is much smaller for the VGG-16 because the network is already trained and hence the initial weights shall only slightly be adapted. Furthermore, the Mini-Batch-Size is much smaller for the training of the VGG-16 since the CNNs are trained on a single GPU with 11 GB RAM.

V. Classification Results

In this section the classification results of the FNN and the CNNs are discussed with respect to their ability for false alarm reduction. At first, the networks are trained with dataset D_1H_1 and tested with D_2H_1 to evaluate the performance for similar data recorded at the same day in the same harbor. The FNN is initialized and trained 99 times, where 80% of the contacts in D_1H_1 are used for training and 20% for validation. Due to higher computational costs the shallow CNN is initialized and trained eleven times. Since primarily only the last layer of the VGG-16 network has to be retrained it is only trained once.

The classification results are illustrated in Fig. 7 in terms of ROC curves. The quality of an ROC curve i.e. the suitability of a classifier is assessed by the area under the ROC curve (AUC). The black ROC curve represents the performance of the feature "Contact SNR" and hence the performance of the standard active signal processing. An increased detector threshold leads on the one hand to less false alarms (i.e. a lower false positive rate (FPR)) and on the other hand to less diver contacts (i.e. a lower true positive rate (TPR)). Moreover, the blue ROC curve illustrates the median performance of the 99 trained FNNs regarding their AUC values. To rate the suitability of a classifier with respect to the ability of false alarm reduction a further criterion is introduced. The false alarm reduction is measured by comparing the FPR of the "Contact SNR" with that of a considered classifier at a TPR of 0.95 which means that 5% of the diver contacts can be discarded. Regarding this, the FPR can be reduced from ≈ 0.36 achieved with the "Contact SNR" to a value of ≈ 0.04 by using the FNN. Thus, the false alarm rate can be reduced by $\approx 89\%$. Furthermore, the result of the median performance of the eleven trained shallow CNNs is represented by the red ROC curve. It can be seen that the performance is much better than that of the "Contact SNR" but also significantly worse than that achieved by the FNN. Finally, the result for the VGG-16 which is represented by the green ROC curve shows that the performance is similar to that of the shallow CNN. For a TPR of 0.95 the false alarm rate can be reduced by $\approx 64\%$ with both CNNs. The classification results are summarized

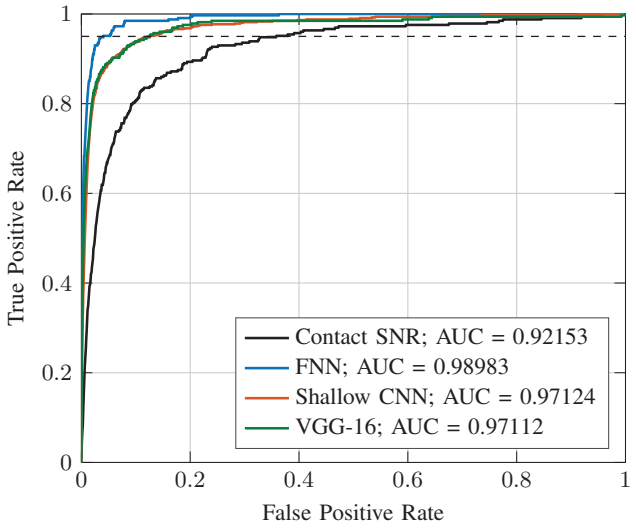


Fig. 7. ROC curves for the FNN, Shallow CNN and VGG-16 when the networks are trained with dataset D_1H_1 and tested with D_2H_1.

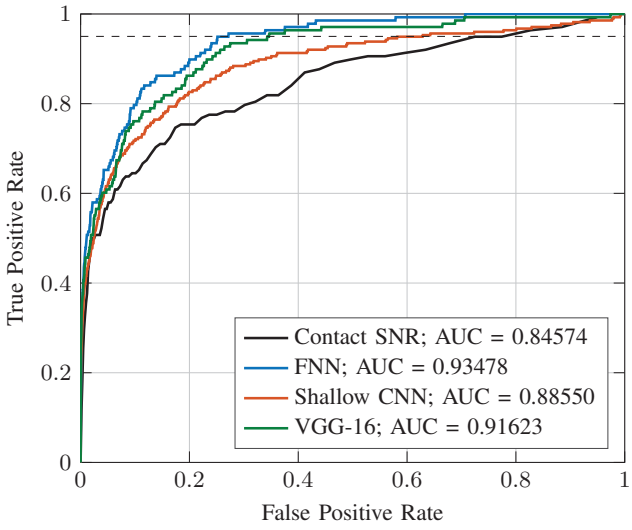


Fig. 8. ROC curves for the FNN, Shallow CNN and VGG-16 when the networks are trained with dataset D_1H_1 and tested with D_3H_2.

TABLE II
COMPARISON OF CLASSIFICATION RESULTS WHEN TRAINING AND TEST DATA ARE SIMILAR

Method	AUC	FPR @ TPR = 0.95	FA Reduction
Contact SNR	0.92153	0.36	–
FNN	0.99019	0.04	89%
Shallow CNN	0.97124	0.13	64%
VGG-16	0.97112	0.13	64%

TABLE III
COMPARISON OF CLASSIFICATION RESULTS WHEN TRAINING AND TEST DATA ARE DIFFERENT

Method	AUC	FPR @ TPR = 0.95	FA Reduction
Contact SNR	0.84574	0.71	–
FNN	0.93478	0.25	65%
Shallow CNN	0.88550	0.58	18%
VGG-16	0.91623	0.34	52%

in Table II which contains the AUC, the FPR for a fixed TPR value of 0.95 as well as the described reduction of the false alarm rate in percent. It can be concluded that the hand-crafted feature based classification method provides the best results in this analysis. The results of the shallow CNN and the VGG-16 are almost the same which suggests that a deeper network performs not necessarily better for this type of data. However, since the use of the CNNs require much less manual engineering the achieved results are surprisingly well.

In a second analysis, the trained classifiers are tested with dataset D_3H_2 to get an impression about the generalization aspect of the classification algorithms. The results are illustrated in Fig. 8. A comparison of the ROC curves of the "Contact SNR" for dataset D_3H_2 and D_2H_1 reveals that the use of the detector threshold leads to a worse performance in dataset D_3H_2. Hence, it can be concluded that the different environmental conditions lead to different statistics in the data which i.e. causes different SNR values. Thus, the input images for the use of the CNNs as well as the extracted features for the use of the FNN might differ from that in dataset D_1H_1 and D_2H_1. However, the use of the FNN as well as the CNNs lead to a significantly better performance than that of the "Contact SNR". It should be mentioned that the results of the FNN and the shallow CNN again show the median performance of all trained networks. For the operational use of a specific FNN or shallow CNN, the networks that achieve the best average performance considering both datasets should be used. The classification results are summarized in Table III. It can be seen that the best results can again be achieved with the use of the hand-crafted feature based FNN. The false alarm rate can be reduced by $\approx 65\%$ for a TPR of 0.95. Furthermore, the performance of the VGG-16 is much better than that of the shallow CNN which indicates that the depth of the network does have an influence on the classification results respectively the generalization to other data.

VI. SUMMARY AND WAY AHEAD

This work proposes a method for false alarm reduction in active sonar applications by means of contact classification. Therefore, two different classification techniques are investigated and its performance is compared.

Firstly, the modification of the standard active signal processing chain is introduced followed by an almost automated method for the labeling of sonar contacts. Furthermore, the inputs and the training steps for the use of the hand-crafted feature based FNN as well as the CNNs are described. Finally, the classification results show that both provided methods outperform the standard active signal processing. Since the trained classifiers show good performances for different datasets it can be concluded that the trained networks are robust. Nevertheless, the best classification results could be achieved with the hand-crafted feature based classification method. However,

it should not be ignored that the training data are highly unbalanced and include only a few diver contacts. Regarding this and the fact that the CNNs do not require a lot of manual engineering it can be concluded that the results are fairly good.

In future work further pre-trained CNNs like the AlexNet or VGG-19 will be analyzed. Furthermore, it will be investigated whether the results achieved with the CNNs can be improved by using more diver contacts in the training set. This will be considered in terms of data augmentation as well as by the use of more datasets. Furthermore, it will be determined if the performance of the hand-crafted feature based classification can be further increased by the use of additional features from the tracking such as the estimated speed or trajectory. Since all classifiers are suitable for the reduction of false alarms it should be analyzed if the false alarm rate can be further reduced by combining several classifiers. Finally, this and further work should result in a robust algorithm which is trained with as many datasets as possible to ensure that it is invariant to different environments.

ACKNOWLEDGMENT

The authors want to thank *ATLAS ELEKTRONIK GmbH* as well as the *Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71)* for the support and funding of this research.

REFERENCES

[1] A. Baldacci, G. Haralabus, S. Coraluppi, and M. Prior, "Adaptive subband processing in active detection and tracking," in *Proceedings of the 2nd International Conference and Exhibition on Underwater Acoustic Measurements: Technologies and Results, IACM - FORTH, Heraklion, Greece*, 2007.

[2] V. Myers, J. Fawcett, P. Hines, and V. Young, "Reconstruction and fusion of perceptual features for automatic classification of sonar echoes," in *OCEANS 2008, Quebec City, Canada*, 2008.

[3] Y. Juan, X. Feng, W. Zhiheng, A. Xudong, L. Jia, J. Yongqiang, and W. Tao, "Adaptive multi-feature fusion for underwater diver classification," in *2013 IEEE/OES Acoustics in Underwater Geosciences Symposium, Rio de Janeiro, Brazil*, 2013.

[4] H. Berg, K. T. Hjelmervik, D. H. S. Stender, and T. S. Sastad, "A comparison of different machine learning algorithms for automatic classification of sonar targets," in *OCEANS 2016, Monterey, California*, 2016.

[5] D. H. S. Stender, H. Berg, K. T. Hjelmeryik, and T. S. Sastad, "The classification performance of signal-to-noise ratio and kinematic features in varying environments," in *OCEANS 2017, Aberdeen, Scotland*, 2017.

[6] M. Buß, S. Benen, D. Stiller, D. Kraus, and A. Kummert, "Feature selection and classification for false alarm reduction on active diver detection sonar data," in *Proceedings of the Underwater Acoustics Conference & Exhibition (UACE) 2017, Skiathos, Greece*, 2017, pp. 569–576.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, Nevada*, 2012, pp. 1097–1105.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

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

[10] ——, "Demystifying deep convolutional neural networks for sonar image classification," in *Proceedings of the Underwater Acoustics Conference, Skiathos, Greece*, 2017, pp. 513–520.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2017.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.

[13] G. Cybenko, *Approximation by Superpositions of a Sigmoidal Function*. Mathematics of Control, Signals, and Systems 2 (1989), 303-314.