# False Alarm Reduction for Active Sonars using Deep Learning Architectures

Matthias Buß[a], Stephan Benen[b], Dieter Kraus[c], Anton Kummert[a]

**Abstract** — In this paper the performance of pre-trained deep convolutional neural networks (CNNs) is compared to that of a classical machine learning approach for false alarm reduction of an active sonar application. Several pre-trained deep CNNs that are firstly introduced in the ImageNet Large Scale Visual Recognition Challenges (ILSVRC) are considered for classifying sonar contacts. The inputs for the CNNs are two-dimensional sonar images (level vs. bearing and time/range) of the contacts. Furthermore, a hand-crafted feature based feedforward neural network (FNN) is considered. 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 adaptable parameter.

## 1 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.

In previous works [1]-[2] 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. Moreover, it could be shown that convolutional neural networks (CNNs) that automatically extract features out of labelled input signals or images are suited for this task.
This work is an extension of the methods described in [2]. The performance of various pre-trained CNNs is compared to that of a hand-crafted feature based feedforward neural network (FNN) regarding their suitability for reducing the false alarm rate with minor degradation of probability of detection. All algorithms are applied to recorded data of active diver detection trials that were carried out in cooperation between the Bundeswehr Technical Center 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 offline 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 in the categories target contact and 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. Ten different CNNs are considered; a shallow CNN trained from scratch and nine pre-trained deep networks that are originally designed for image classification (AlexNet [3], GoogLeNet [4], Inception v3 [5], ResNet-18, ResNet-50, ResNet-101 [6], SqueezeNet [7], VGG-16 and VGG-19 [8]). The classification performance is assessed by means of Receiver-Operating-Characteristic (ROC) curves and the generalisation respectively the robustness of the classifiers is proved by testing the algorithms with unseen data recorded in different environments.

## 2 Data for Classification

The contacts that are forwarded to the machine learning algorithms are derived by a standard active signal processing chain that contains the steps beamforming, matched-filtering, normalisation, detection and tracking. The results from the tracking are used as ground truth for contact labelling. All contacts that belong to the track that is assigned to the diver, are labelled as “Diver Contact” and all other contacts are labelled as “False Alarm”. A more detailed description of the labelling process can be found in [2]. In total six datasets from three different diver detection trials that are recorded in different environments are considered. From each of the three trials two datasets are chosen; one for training and one for testing the classifi-

[a] University of Wuppertal, {matthias.buss ; kummert}@uni-wuppertal.de
[b] ATLAS ELEKTRONIK GmbH, stephan.benen@atlas-elektronik.com
[c] City University of Applied Sciences Bremen, dieter.kraus@hs-bremen.de

**Table 1**: Information for considered datasets for classification.

| **Shortcut** | **Pulse Parameters** | | | | **# Diver Contacts** | **# False Alarms** |
|---|---|---|---|---|---|---|
| | ***Type*** | ***Centre Frequency*** | ***Bandwidth*** | ***Length*** | | |
| $D_{\text{Train}}E_1$ | HFM | 100 kHz | 12 kHz | 50 ms | 255 | 21831 |
| $D_{\text{Train}}E_2$ | LFM | 100 kHz | 20 kHz | 100 ms | 136 | 21141 |
| $D_{\text{Train}}E_3$ | HFM | 100 kHz | 20 kHz | 100 ms | 320 | 3761 |
| $D_{\text{Test}}E_1$ | HFM | 100 kHz | 12 kHz | 50 ms | 356 | 37843 |
| $D_{\text{Test}}E_2$ | LFM | 100 kHz | 20 kHz | 100 ms | 194 | 22484 |
| $D_{\text{Test}}E_3$ | HFM | 100 kHz | 20 kHz | 100 ms | 187 | 2484 |

cation algorithms. Some information about the considered datasets are given in Table 1. The expressions $E_1$, $E_2$, $E_3$ in the shortcut indicate the three different environments in which the data were recorded. Since the performance of CNNs increases with the amount of training data, the three listed training datasets from different diver detection trials ($D_{\text{Train}}E_1$, $D_{\text{Train}}E_2$, $D_{\text{Train}}E_3$) are merged to a big training dataset. Furthermore, three other datasets from the same trials as the three training datasets are chosen as test datasets ($D_{\text{Test}}E_1$, $D_{\text{Test}}E_2$, $D_{\text{Test}}E_3$). From Table 1 it can be seen that the pulse parameters as well as the number of diver contacts and false alarms differ from dataset to dataset.

## 3 Classification Methods

In this work two different classification techniques are considered. On the one hand a classical hand-crafted feature based technique and on the other hand convolutional neural networks that intrinsically extract relevant features of given input images or signals during the classification process. All considered neural networks are trained and tested with the Neural Network Toolbox in MATLAB.

### 3.1 Hand-Crafted feature based classification

The hand-crafted feature based classification is considered by means of an FNN. The active signal processing chain is extended by a feature extraction method. For each contact resulting from the detection, in total 53 features from different categories are extracted. Some examples for the extracted features are given with respect to the snapshot of the two-dimensional data of a diver contact presented in Figure 1. 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. The lower image 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 ≈ 1 m and the extent in beam direction is 5 Beams. The upper image illustrates the corresponding normalised data, in which the contact pixels 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 exceeds 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”.

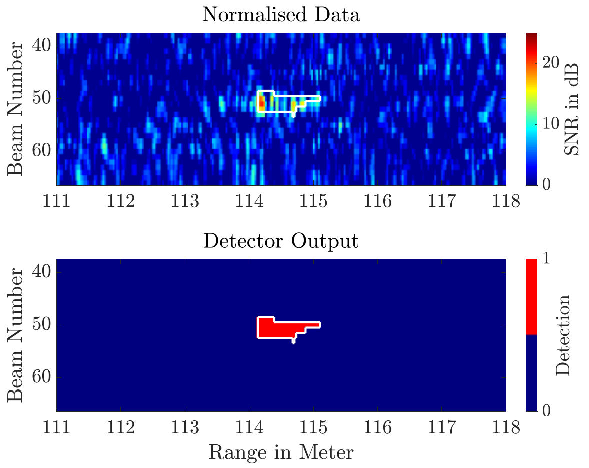

**Figure 1**: Exemplary diver contact after normalisation and detection.

The extracted features are used to train a feedforward neural network (FNN). According to the universal approximation theorem, an 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 bounded, continuous and nonconstant [9]. Regarding this, the FNN in this work consists of one hidden layer and the activation function is the hyperbolic tangent which fulfils the aforementioned constraints. The structure of the used FNN is illustrated in Figure 2. It can be seen that the input of the network consists of 53 input variables representing the extracted features. These are forwarded to 20 neurons in the hidden layer. The output layer of the network consists of two neurons and uses the softmax function for the calculation of a probability of class affiliation.

The final weights of an FNN that result from the training process depend on their initial setting. Therefore, in this work the FNN is trained 30 times with different random initialisations and the best performing constellation is selected on the basis of an appropriate performance criterion that will be described later.

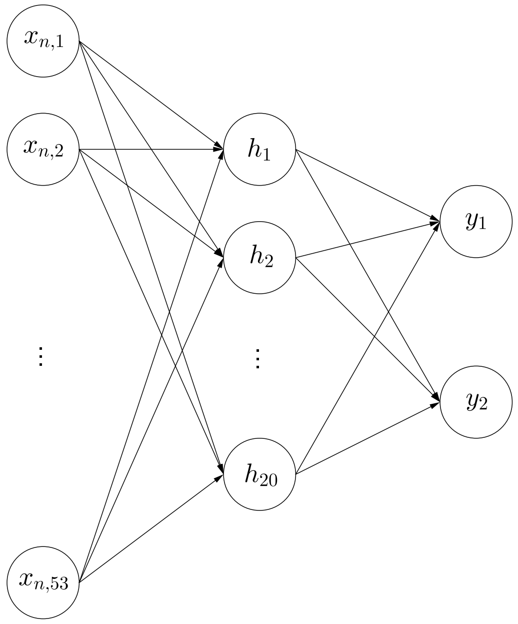

**Figure 2**: Structure of used Feedforward Neural Network.

### 3.2 Convolutional Neural Networks

In addition to the hand-crafted feature based FNN, ten different CNNs are considered. These require a two-dimensional image as input. Hence, for each contact a region of interest (ROI) of the normalised two-dimensional sonar images is extracted. An example for the extraction of the input data can be given by the diver contact shown in Figure 1. The normalised 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 Figure 3.

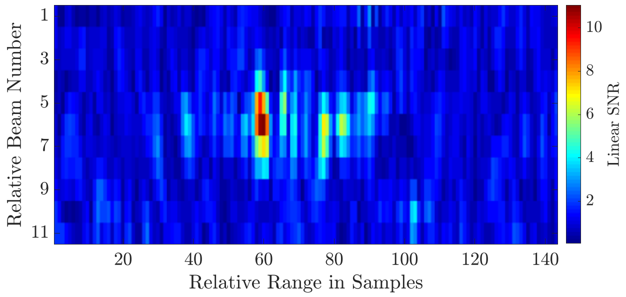

**Figure 3**: Exemplary ROI that is used as input image for CNNs.

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 normalised data. In this work on the one hand a shallow CNN which is trained from scratch is considered. The structure of the network is illustrated in Figure 4. It can be seen that the network consists of only one convolutional layer followed by an average pooling layer. Furthermore, the fully connected layer consists of one hidden layer with 4096 neurons and one output layer with two neurons for binary classification.

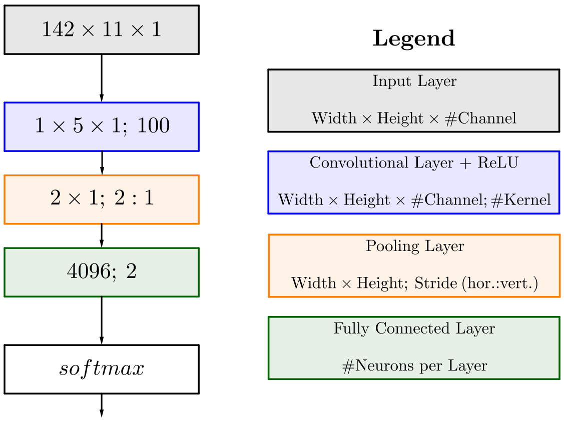

**Figure 4**: Structure of CNN trained from scratch.

On the other hand, the pre-trained deep networks AlexNet, GoogLeNet, Inception v3, ResNet-18, ResNet-50, ResNet-101, SqueezeNet, VGG-16 and VGG-19 are considered. These networks are firstly introduced in the ImageNet Large Scale Visual Recognition Challenges (ILSVRC) and originally trained to distinguish images of 1000 different classes containing various types of ships, animals, cars, food, etc. However, in this work their performance for classifying sonar contacts is considered. Since the pre-trained networks require images that have a size of

- 224 × 224 × 3 (GoogLeNet, ResNet, VGG)
- 227 × 227 × 3 (AlexNet, SqueezeNet)
- 299 × 299 × 3 (Inception)

it is necessary to adapt the extracted ROIs of the sonar contacts. Therefore, the ROIs of size 142 × 11 have to be resampled to the above-mentioned size. It can be seen that the third dimension of the input size is three which means that all networks use R-G-B images as input. In this work the ROIs of the sonar contacts are stored as grey scale images that are then forwarded to all three channels (R, G and B). In addition to the adaptation of the input images also the output layers of the networks have to be modified. Since the pre-trained networks are trained for distinguishing images of 1000 different classes, the output layers with 1000 neurons have to be replaced by output layers with two neurons for distinguishing diver contacts and false alarms.

For the transfer learning of the pre-trained networks the training options in MATLAB are set to

- `sgdm` (stochastic gradient descent with momentum)
- `MaxEpochs: 6`
- `MiniBatchSize: 50`
- `InitialLearnRate: {1e-4, 1e-3, 5e-2, 1e-2}`
- `Shuffle: every-epoch`
- `LearnRateSchedule: piecewise`
- `LearnRateDropPeriod: 2`
- `LearnRateDropFactor: 1e-1.`

It can be seen that four different initial learning rates are considered. The aim of this is to investigate whether a slight or a strong adjustment of the kernel weights leads to better classification results. Since the output layer is replaced and the weights are initialised randomly, the learning rate for the output layer may need to be higher than the learning rate for the kernel weights in the convolutional layers. Therefore, the learning rates mentioned above can be multiplied by the factor given with the training options

- `WeightLearnRateFactor`
- `BiasLearnRateFactor`

to increase the learning rates for the output layer. Four different weighting factors (1, 10, 50, 100) are considered, so that in total 16 different constellations of each network are trained. Moreover, each network is trained three times which results in 48 different constellations.

## 4 Classification Results

All previously described networks are trained with the merged training dataset consisting of $D_{\text{Train}}E_1$, $D_{\text{Train}}E_2$ and $D_{\text{Train}}E_3$ and tested with each of the three test datasets. The performance of the classifiers is measured by means of Receiver-Operating-Characteristic (ROC) curves. In Figure 5 exemplarily four different ROC curves are shown. The performance of the standard active signal processing for dataset $D_{\text{Test}}E_2$ is displayed by the black ROC curve. The false positive rate (FPR) of one means that 100% of the 22484 detected false alarms are available. Similarly, the true positive rate (TPR) of one indicates that all 194 diver contacts are present. An increase of the detection threshold in the standard signal processing leads on the one hand to a lower FPR but on the other hand to a lower TPR. In the ideal case the TPR stays at a value of one (all diver contacts remain) whereas the FPR goes to zero (no false alarms).

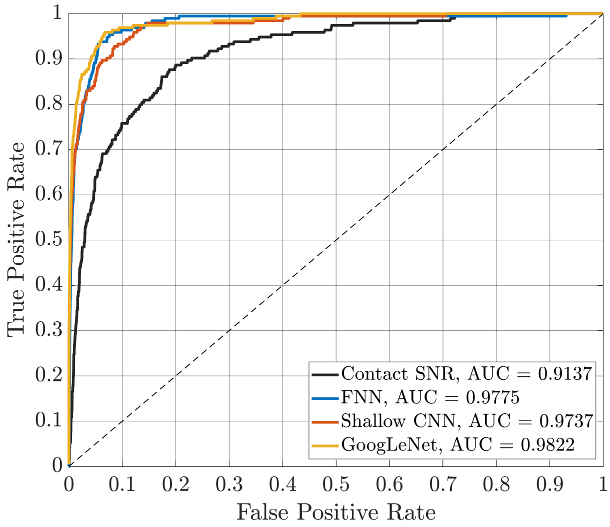

**Figure 5**: ROC curves of the Contact SNR, FNN, Shallow CNN and GoogLeNet for test dataset $D_{\text{Test}}E_2$.

In addition to the ROC curve for the Contact SNR also the results of the hand-crafted feature based FNN as well as the shallow CNN trained from scratch and the transfer learned GoogLeNet are displayed. It can be seen that all classifiers achieve a significantly lower FPR compared to the Contact SNR for almost all TPR values. Furthermore, the ROC curves illustrate that none of the three considered classifiers performs best in all areas.

Since in total 30 constellations of the FNN and 48 of each pre-trained CNN are trained, the best performing networks have to be figured out by a suitable performance criterion. In this work the area under the ROC curve (AUC) is chosen. For each trained network the AUCs for the three test datasets are calculated and finally the network that leads to the highest mean AUC over all three test datasets is selected for further analyses. For all pre-trained CNNs different combinations of the training parameters “InitialLearnRate”, “WeightLearnRateFactor” and “BiasLearnRateFactor” lead to the best performing constellations so that no general statement for the choice of this parameters can be made.

The results of the best performing networks are illustrated in Figure 6 for all test datasets. Furthermore, the mean AUC over all three datasets is displayed by the violet coloured bars. It can be seen that the classification results of all networks lead to much higher AUCs than the AUC achieved with the standard active signal processing represented by the “Contact SNR”. This means that the hand-crafted feature based classification with the FNN as well as the CNNs outperform the standard signal processing significantly. On closer inspection it can be seen that none of the classifiers performs best in all three datasets. It seems that all convolutional neural networks are overfitted to the first two training datasets since they perform similarly good for the first two test datasets and worst for the third test dataset. The most likely explanation for this is the amount of training data in the third training dataset which is approximately five times less than in the two other training datasets. However, a high fluctuation of the performances for the three test datasets does not appear with the hand-crafted feature based FNN. Another important outcome of this evaluation can be concluded by comparing the results of the shallow CNN and the deep CNNs. The mean performance of all pre-trained deep CNNs is higher than that of the shallow CNN which indicates that the depth of the network has an influence on the performance.

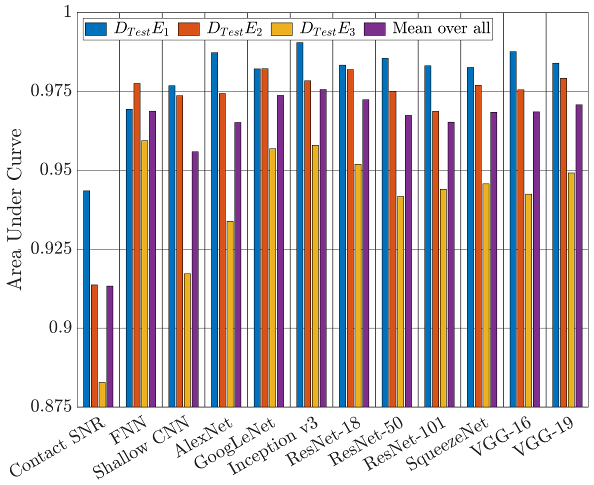

**Figure 6**: AUC values of the standard signal processing and all considered neural networks for each test dataset.

Since the main focus of this work is on the false alarm reduction, a second criterion for assessing this is introduced. The FPR achieved by the different classification algorithms is compared to that obtained by the Contact SNR at a TPR of 0.9. As an example, the ROC curves displayed in Figure 5 show that using the Contact SNR an FPR of 0.23 can be achieved at a TPR of 0.9. With the FNN an FPR of only 0.05 can be achieved at the same TPR resulting in a false alarm reduction of 78% for dataset $D_{\text{Test}}E_2$. In Figure 7 the results of the false alarm reduction criterion are illustrated for all considered classification algorithms and test datasets. It can be seen that the classification with the FNN leads to an average false alarm reduction of 75%. Moreover, the classification with the shallow CNN reaches 58% whereas all deeper CNNs perform much better. Regarding this criterion, the best performance can be achieved with the VGG-19, with that on average only 18% of the false alarms remain.

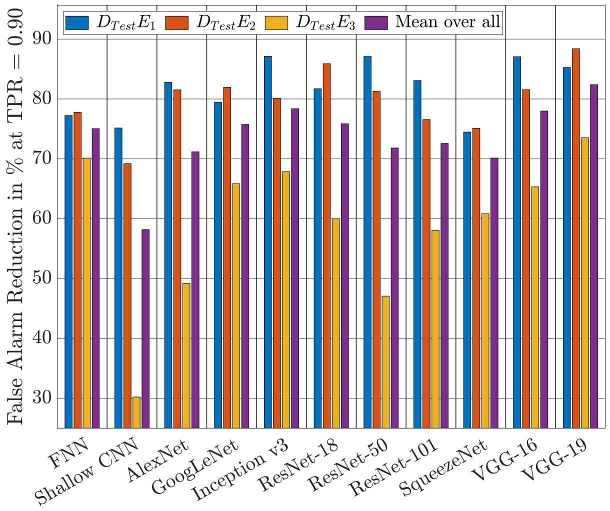

**Figure 7**: False Alarm Reduction at a true positive rate of 0.9 for each test dataset and each considered classifier.

## 5 Summary and Way Ahead

In this paper, a classical machine learning technique based on hand-crafted features as well as pre-trained CNNs are investigated for classifying sonar targets. Even though the pre-trained CNNs are designed for classification in R-G-B images, the performance on the sonar data that only contain SNR levels is quite good. A comparison of the performance of the shallow CNN trained from scratch with that of the transfer learned deep CNNs demonstrates that the deeper networks perform better. Since some of the CNNs perform slightly better than the hand-crafted feature based FNN, it can be concluded that the hand-crafted features do not address all possible relevant information for distinguishing target contacts and false alarms.

In future work on the one hand, the extraction of further suitable hand-crafted features should be considered. On the other hand, the combination of CNNs with hand-crafted features should be analysed. This could either be done by combining the final feature map of a CNN with the hand-crafted features that are further used to train another FNN or by feeding the hand-crafted features as additional inputs, that are forwarded to the fully connected layer, into a CNN. Furthermore, a deep CNN should especially be designed for sonar data.

## Acknowledgement

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] M. Buß et al., Feature selection and classification for false alarm reduction on active diver detection sonar data, UACE 2017, pp. 569–576 (2017)

[2] M. Buß et al., Hand-Crafted Feature Based Classification against Convolutional Neural Networks for False Alarm Reduction on Active Diver Detection Sonar Data, Oceans 2018, pp. 1-7 (2018)

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, NIPS 2012, pp. 1097–1105 (2012)

[4] C. Szegedy et al., Going deeper with convolutions, CVPR 2015, pp. 1-9 (2015)

[5] C. Szegedy et al., Rethinking the Inception Architecture for Computer Vision, CVPR 2016, pp. 2818-2826 (2016)

[6] K. He et al., Deep Residual Learning for Image Recognition, CVPR 2016, pp. 770-778 (2016).

[7] F. N. Iandola et al., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size (2016)

[8] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, ICLR 2015, (2015)

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