A Lightweight CNN for Noise Source Detection in Bearing-Time Sonar images

Soggia Antonio, Lionel Fillatre, Deruaz-Pepin Laurent

A Lightweight CNN for Noise Source Detection in Bearing-Time Sonar images

1st Soggia Antonio
Thales DMS, I3S
Sophia-Antipolis, France
antonio.soggia@etu.univ-cotedazur.fr

2nd Fillatre Lionel
Université Côte d'Azur, CNRS, I3S
Sophia-Antipolis, France
lionel.fillatre@i3s.unice.fr

3rd Deruaz-Pepin Laurent
Thales DMS
Sophia-Antipolis, France
Laurent.Deruaz-Pepin@fr.thalesgroup.com

***Abstract*—In current broadband passive sonar systems, a human operator observes a series of sonar measurements that represents different energy levels in the bearing and time dimensions. Detecting weak noise sources amidst ambient sea noise can be challenging. This paper proposes a lightweight neural network based on U-Net that detects the presence of low SNR noise sources in a multipath setting. The architecture we designed predict only the direct path. Then, a linear detector can detect efficiently the presence of the acoustic source from the enhanced direct path.**

***Index Terms*—CNN, detection, contrast enhancement, sonar, multipath beamforming**

I. Introduction

Passive sonar systems are used for detecting underwater objects without emitting sound. They display broadband signals to the operator as a 2D (bearing-time) image like the noised one in Figure 1. In [1] we discussed the drawbacks

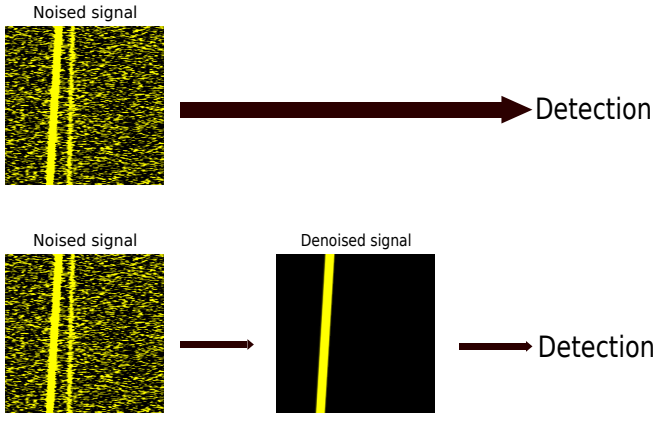

Fig. 1. Comparison between a traditional detection pipeline (up) and ours with image enhancement before the pixel-wise detection (down).

of passive sonar systems, mainly the inability to discriminate source signal from different noise sources and the unreliability in detecting attenuated signals. To solve these problems we tried conventional contrast-enhancing algorithms [2], [3], [4], [5] that must be calibrated and that produce artefact and phantom sources. Moreover, wavelets [6], [7] have also been explored for contrast-enhancing. However, while they offer certain advantages, they come with their own drawbacks [8]. Wavelets suffer from artifacts such as over-smoothing and computational complexity. We also tried a gradient-based detection algorithm approach using Canny [9] and LSD [10]. They performed poorly in presence of strong noise. The noise in the sonar images is too complex for this kind of algorithms as it presents high-frequency and Gaussian components. It makes the noise be falsely detected as a signal. Our previous paper used a small and reliable Unet-based architecture to enhance the noise-signal contrast. It proved well but its effects were limited: it didn't account for signal attenuation in time [11], [12] and signal reflection [13], [14].

In this paper, we update the data generation process to account for more complexity. The main contributions of this work are manifold. First, we improve our previous sonar image formation model by incorporating signal attenuation and reflection. Multipath causes ghost targets to appear dynamically in sonar images. This new physical model is used to train our lightweight CNN (Convolutional Neural Network), called AntoNet, such that it can remove multipath interferences in sonar images. Secondly, we compare several versions - with a different number of parameters - of our linear lightweight CNN with other state-of-the-art ones ([15], [16], [17]). Thirdly, after removing multipath interferences, we detect the source signal with an almost optimal detector. We show that AntoNet improves significantly the detection results. The reason as to why we use a pixel-wise detector stems in sonar image processing. It needs to be precise and to remove as much false alarm as possible. Another reason is that CNN operate in a local pixel setting [18]. ROC curves detection values are directly derived from the simulation model. This makes the detection assessment accurate and adaptable to complex sonar images.

This paper is organized as follows. Section II describes the enhanced sonar image formation model, including detailed simulation of acoustic propagation and reflection phenomena. Section III presents the AntoNet architecture and the associated contrast enhancement methodology. In Section IV, we provide extensive numerical results and detection evaluations. Section V concludes the paper.

II. Sonar image formation

Obtaining sonar images in bearing-time like the first image in Fig. 1 is very difficult because there are many industrial

Thanks to ANRT agency for funding.

constraints and the ground truth generally doesn't exist. In order to use high-quantity of training images and to have a better grasp of the data, we have opted to simulate our signals and the noise.

A. Source and antenna model

We assume that the source is unique and follows a uniform rectilinear movement over the sea. This source is emitting a monochromatic wave with an unknown frequency ν. However, we account for its reflections against the surface and the seabed. We also account for its attenuation in relation to the distance. We assume that the surfaces are not flat but irregular, as shown in Fig. 2. This makes the sound reflections scatter in all direction, with random amplitude [14]. Because of this, we can assume that the three paths are independent.

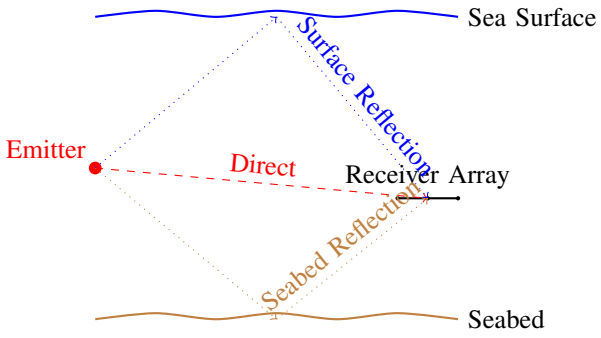

Fig. 2. Scheme of sound propagation showing the emitter, the receiver array, and the reflection paths from the sea surface and seabed.

We define three paths; each is associated to a letter $\ell \in \{d, s, b\}$ that represents the index, respectively, for direct, reflected to the surface and to the seabed trajectory. The signal is measured with a linear array. Then, for each ℓ, we define the steering vector $\boldsymbol{h}_t^{(\ell)}$ at time t as:

$$\boldsymbol{h}_t^{(\ell)} = \left[1, \dots, e^{2i\pi\nu\delta_k\left(\theta_t^{(\ell)}\right)}, \dots, e^{2i\pi\nu\delta_K\left(\theta_t^{(\ell)}\right)}\right], \quad (1)$$

with

$$\delta_k(\theta_t) = (k-1)\frac{d\cos(\theta_t)}{c}, \quad (2)$$

where d is the constant distance between sensors, c is the wavefront speed and θ is the source angle projection onto the antenna. Different directions have different θ.

We define the total distances traveled by the waves as L_d, L_s and L_b. We consider a (x, y, z)-coordinate system where $(0, 0, z_0)$ is the receiver position. The depth z is relative to the surface and Z is the seabed depth. A short calculation yields:

$$L_d = \sqrt{x^2 + y^2 + (z - z_0)^2} \quad (3)$$

$$L_s = \sqrt{x^2 + y^2 + (z + z_0)^2} \quad (4)$$

$$L_b = \sqrt{x^2 + y^2 + (z + z_0 - 2Z)^2}. \quad (5)$$

Each wave is attenuated by its own attenuation factor $\alpha_d = \frac{1}{L_d}$, $\alpha_s = \frac{A_s}{L_s}$ and $\alpha_b = \frac{A_b}{L_b}$, where A_s and A_b are the absorption factors of the indirect trajectories.

B. Measurement model

At time t, we assume that we have N stationary sonar measurements $\boldsymbol{s}_{t,i} \in \mathbb{C}^K$, for $i = 1, \dots, N$, given by

$$\boldsymbol{s}_{t,i} = a_{t,i}^{(d)}\alpha_d\,\boldsymbol{h}_t^{(d)} + a_{t,i}^{(s)}\alpha_s\,\boldsymbol{h}_t^{(s)} + a_{t,i}^{(b)}\alpha_b\,\boldsymbol{h}_t^{(b)} + \boldsymbol{\eta}_{t,i}, \quad (6)$$

where $\boldsymbol{\eta}_{t,i}$ is a sequence of independent circularly-symmetric central complex normal noise with zero mean and covariance matrix $\sigma_\eta^2 I_K$ and I_K is the identity matrix of size K. The signal amplitudes $a_{t,i}^{(\ell)}$ are independent sequence of complex normal variable with zero mean and variance σ_a^2 for each source, due to the reflections on rough surfaces. Hence, the sonar measurements are processed as a batch of N vectors of size K. The SNR, measured in dB, is defined by

$$\text{SNR} = 10\log_{10}\left(\varrho^2\sqrt{N}\right), \text{with } \varrho^2 = \frac{\sigma_a^2}{K\sigma_\eta^2}, \quad (7)$$

where ϱ^2 is the peak-SNR.

Let

$$\boldsymbol{g}_j^+ = [1, \dots, e^{-i\pi k\frac{j}{K-1}}, \dots, e^{-i\pi(K-1)\frac{j}{K-1}}] \quad (8)$$

be the steering vector for the jth channel beamforming when $j = 0, \dots, K-1$. The integrated energy $x_{t,j}$ on the jth channel at time t is computed by

$$x_{t,j} = \frac{1}{N}\sum_{i=1}^{N}\left|\boldsymbol{g}_j^+\boldsymbol{s}_{t,i}\right|^2, \quad (9)$$

where $\boldsymbol{g}^+$ denotes the conjugate transpose of the vector $\boldsymbol{g}$. Let us define $G_{t,j}^{(\ell)}$ for each $\ell \in \{d, s, b\}$ as

$$G_{t,j}^{(\ell)} = \boldsymbol{g}_j^+\boldsymbol{h}_t^{(\ell)} = \sum_{k=1}^{K} e^{-i\pi\frac{j}{K-1}(k-1)}\,e^{2i\pi\nu\,\delta_k\left(\theta_t^{(\ell)}\right)}. \quad (10)$$

We finally get

$$x_{t,j} = \frac{1}{N}\sum_{i=1}^{N}\left|\sum_{\ell\in\{d,s,b\}} a_{t,i}^{(\ell)}\,\alpha_\ell\,G_{t,j}^{(\ell)} + \boldsymbol{g}_j^+\boldsymbol{\eta}_{t,i}\right|^2. \quad (11)$$

Let us denote $\mathcal{H}_0(t, j)$, resp. $\mathcal{H}_1(t, j)$, the case when the acoustic source is not present, resp. is present, in the channel j at time t. After an appropriate standardization of the channel (we do not change the notation to keep it simple), we get the following asymptotic distribution

$$x_{t,j} \sim_{N\to\infty} \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0(t,j), \\ \mathcal{N}\left(\varrho_{t,j}^2\sqrt{N}, \left(1+\varrho_{t,j}^2\right)^2\right) & \text{under } \mathcal{H}_1(t,j), \end{cases} \quad (12)$$

with

$$\varrho_{t,j}^2 = \frac{\sigma_a^2}{K\sigma_\eta^2}\left(|\alpha_d\,G_{t,j}^{(d)}|^2 + |\alpha_s\,G_{t,j}^{(s)}|^2 + |\alpha_b\,G_{t,j}^{(b)}|^2\right). \quad (13)$$

The derivation of this approximation is based on the fact that the reflected signals are independent from the direct one and that the amplitudes are i.i.d.

The sonar image is then defined as the matrix

$$X = (x_{t,j})_{1\le t\le T, 0\le j\le K-1} \in \mathbb{R}^{T\times K} \quad (14)$$

where T is the last measurement time. Since we want to remove the reflected signals and the noise, we want to predict the SNR image given by

$$Y = (y_{t,j})_{1\leq t\leq T, 0\leq j\leq K-1} \in \mathbb{R}^{T\times K} \tag{15}$$

with

$$y_{t,j} = \begin{cases} 0 & \text{under } \mathcal{H}_0(t,j), \\ \tilde{\varrho}^2_{t,j}\sqrt{N} & \text{under } \mathcal{H}_1(t,j), \end{cases} \tag{16}$$

where $\tilde{\varrho}^2_{t,j}$ is the SNR of the direct trajectory:

$$\tilde{\varrho}^2_{t,j} = \frac{\sigma_a^2}{K\sigma_\eta^2}|G^{(d)}_{t,j}|^2. \tag{17}$$

Fig. 3 shows the noisy sonar image and the labeled image with a -2dB signal.

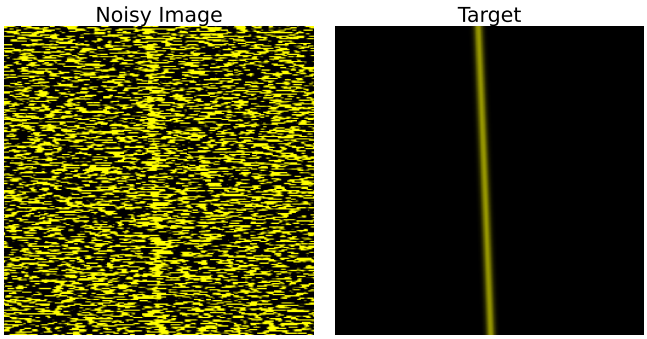

Fig. 3. An example of generated sonar image X (left) and its associated label image Y (right) for a -2dB SNR source signal.

C. Pre-processing

To help the CNN with the learning, we apply simple contrast enhancing algorithms to all the images. For each pixel $x_{t,j}$, we first apply $\mathcal{C} : \mathbb{R} \to \mathbb{R}$ such that

$$\mathcal{C}(x) = \begin{cases} \alpha \cdot \xi(x), & \text{if } \xi(x) \leq \gamma \\ 1 + \alpha \cdot \gamma \cdot \ln(\xi(x)), & \text{if } \xi(x) > \gamma \end{cases} \tag{18}$$

$$\xi(x)=\frac{x - x_{\min}}{x_{\max} - x_{\min}},\ \alpha=\frac{1}{\gamma \cdot (1 - \ln(\gamma))},\ \gamma=\frac{x_{\text{threshold}} - x_{\min}}{x_{\max} - x_{\min}},$$

and $x_{\text{threshold}}$ is chosen such that we can control where to scale the high values. This function helps to scale stronger values while retaining unchanged weakest ones. The function $\mathcal{C}$ is invertible ; its inverse is denoted $\mathcal{C}^{-1}$.

We then define the preprocessing function $\psi : \mathbb{R} \to \mathbb{R}$ to scale higher noise values and remove negative values:

$$\psi(x_{t,j}) = \begin{cases} \log(a + \mathcal{C}(x_{t,j})) & \text{if } \ \mathcal{C}(x_{t,j}) \geq b, \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

with $a = \mathcal{C}(4)$ and $b = \mathcal{C}(-3)$. This function would help the CNN without removing significant information. It removes only extremely negative noise values. We have chosen $b = \mathcal{C}(-3)$ to remove noise values and $a = \mathcal{C}(4)$ to avoid already negative/zero values with the logarithm. The function ψ is not bijective. To make the output comparable to the input X we define ψ^- such that:

$$\psi^-(u) = \mathcal{C}^{-1}(e^u - a). \tag{20}$$

Our neural network is designed to approximate the preprocessed label $\psi(Y)$ for the input X where the notation $\psi(Y)$ means that ψ is applied to each pixel of Y. The post-processing ψ^- allows us to revert the estimated label computed by our neural network to the original noised signal domain.

III. Contrast enhancement with a CNN

We suppose to have a data set consisting of N sonar image pairs $(X^{(n)}, Y^{(n)})$ where $X^{(n)}$ is given in (14) and $Y^{(n)}$ is defined in (15). We choose, to estimate the source power, the empirical mean square error (MSE) as the loss function:

$$\widehat{\text{MSE}}(f)=\frac{1}{NTK}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{K-1}\left(f(\psi(x^{(n)}_{t,j})))-\psi(y^{(n)}_{t,j})\right)^2. \tag{21}$$

To enhance the contrast in the sonar images, we will compare different versions of AntoNet [1] which is a small and interpretable variant of Unet [19], calibrated to better suit our problem. While other networks inspired by Unet, such as C-Unet [15] or DeepUnet [20] are similar to ours in architecture, they are better suited for classification or segmentation. They also contain non-linearities, making them difficult to interpret. AntoNet structure is shown in Fig. 4. It belongs to the family of encoder-decoder models: its structure is composed of two parts, the encoding and the decoding. Concretely, we have 4 encoding and 4 decoding layers. Each encoding component is formed by one convolution followed by an Average Pooling. The decoding layers have transpose convolutions followed by a classic one. After many attempts to keep the model as simple as possible, we decided to make it such that each convolution layer has 3 feature maps. We refer to it as the "classic" AntoNet. We explore also how the parameter number and the depth of the model play a role in performance detection. It is also relevant to compare it to C-Unet, a shallow model but with a high-parameter count, as discussed in section IV.

These architectures are trained on a training dataset consisting on 20,000 sonar images of size 256x256 pixels with SNR varying between -10 dB and 10 dB. They are tested on 2,000 test sonar images with a given SNR of -2 dB and same size of the trained ones. Since Unet and Orca have very similar architectures, their differences are minimal when we adapt them to a contrast-enhancement task. To make the comparison interesting, we use a Unet variant that has 16 Feature Maps (FM) in the first layer instead of 64. We refer to it as Unet-16.

The test is done on C-Unet, Unet-16, ORCA and different AntoNet configurations. In fact, it is interesting to see how much the parameter number and the model depth influences the result. We therefore propose 4 AntoNet settings: the original architecture, a deeper one, a 1 FM per layer version and a 10 FM per layer one. We evaluate our approach with different metrics: training time, inference time and ROC curves. Note that the non-linearities have already been explored in [1] and it

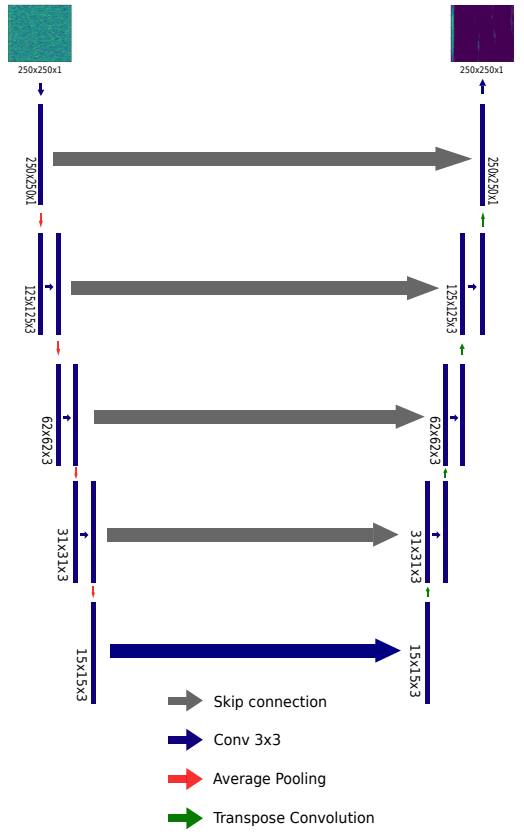

Fig. 4. AntoNet structure: 4 encoding layers consisting of one convolution followed by a pooling operation each. They are followed by 4 decoding layers separated by oversampling and concatenated with the encoded layers of the same dimension.

was shown that they didn't significantly influence the contrast-enhancement performance.

Tables I and II compare the models in terms of the number of parameters, the memory they occupy and the training epoch/inference time. Our model, running in a laptop RTX

TABLE I
COMPARISON OF MODEL SIZES, TRAINING AND INFERENCE TIME

Model	C-Unet	Unet-16	ORCA/UNet
#Params	213k	1.94M	57.14M
Training (epoch)	58s	350s	724s
Testing	5s	40s	58s

TABLE II
COMPARISON OF MODEL SIZES, TRAINING AND INFERENCE TIME

AntoNet	Classic	1 FM	10 FM	Deep
#Params	1.19k	148	12.6k	1.5k
Training	7s	3s	15s	9s
Testing	900ms	500ms	3s	1.48s

4090 GPU, takes less than one second to test 2,000 images while Unet-16, resp. ORCA, takes 40, resp. 58, seconds. Applied on the same image, the different AntoNet settings give similar visual contrast enhancement as shown in Fig. 5.

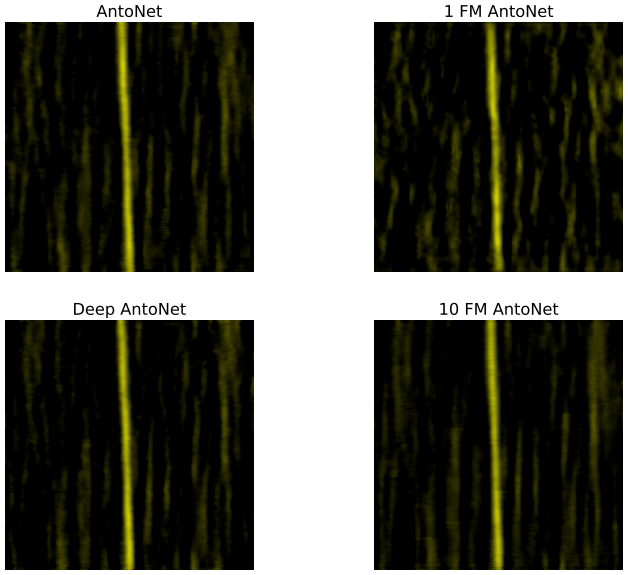

Fig. 5. All models predictions, we can see a small trace of the reflected source to the right that was invisible in the noised image due to the low signal power. The label is shown in Figure 3.

IV. DETECTION PERFORMANCE RESULTS

We evaluate AntoNet performance for detecting signal pixels with a dedicated detector. The performance of a detector assuming a binary decision hypothesis test is characterized by ROC curves. In our case, the test is derived from the hypotheses (12). After applying AntoNet or any other algorithm, we suppose that the reflected signals have been removed. Hence, hypothesis $\mathcal{H}_0 = \mathcal{H}_0(t,j)$ corresponds to noise only or to a removed reflected signal in pixel (t,j). In other words, we assimilate reflected signals to background noise after the preprocessing. If we decide $\mathcal{H}_1 = \mathcal{H}_1(t,j)$ for a pixel that initially contains a reflected signal, it is considered as a false alarm. Hence, we want to decide between the two following hypotheses:

$$x \sim \mathcal{N}(0,1) \quad \text{under} \quad \mathcal{H}_0, \tag{22}$$

$$x \sim \mathcal{N}\left(\varrho^2\sqrt{N}, \left(1+\varrho^2\right)^2\right) \quad \text{under} \quad \mathcal{H}_1, \tag{23}$$

where $\varrho > 0$ is known. A short calculation (ignoring some constant terms) yields

$$\delta^*(x) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda^*(x) = x + \left(\frac{1+\frac{\varrho^2}{2}}{\sqrt{N}}\right) x^2 < \lambda, \\ \mathcal{H}_1 & \text{otherwise.} \end{cases} \tag{24}$$

Assuming that N is large enough and that ϱ^2 is low enough, we can approximate $\delta^*(x)$ by the linear test

$$\delta^*_{\text{lin}}(x) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda^*_{\text{lin}}(x) = x < \lambda, \\ \mathcal{H}_1 & \text{otherwise.} \end{cases} \tag{25}$$

This gives us the probabilities of detection and false alarm, P_d and P_{fa} in function of ϱ^2:

$$P_{fa} = \mathbb{P}(x > \lambda | \mathcal{H}_0) = 1 - \Phi(\lambda), \tag{26}$$

$$P_d(\varrho^2) = 1 - \Phi\left(\frac{\lambda - \varrho^2\sqrt{N}}{1+\varrho^2}\right), \tag{27}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

ROC curves in Fig. 6 show different versions of our model detection performance for signals with a -2dB SNR. We can see here that more parameters yields better performance for the linear model, while a deeper version isn't significantly better. They all however perform relatively well against a noisy input. In Fig. 7 we took our original model and compared its

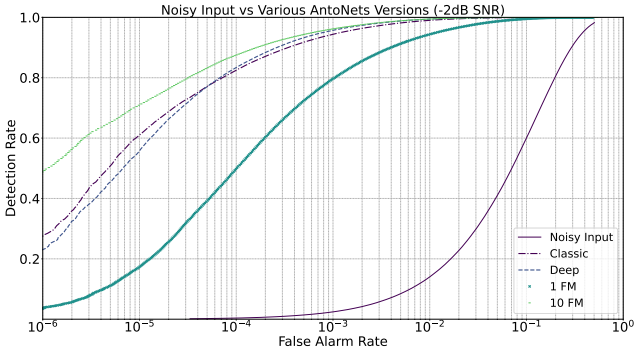

Fig. 6. ROC curves comparing the noisy input with different AntoNet output using a pixel-wise detector performance.

detection performance with different models. It outperforms the others by a large margin while retaining a low parameter count. Linear models, compared to nonlinear large ones (Orca, UNet-16) behave differently. In fact, while ours are well suited for these kind of tasks, bigger models apply a much more aggressive cleaning. In fact, they produce almost zero false alarms but they remove low-SNR signals too.

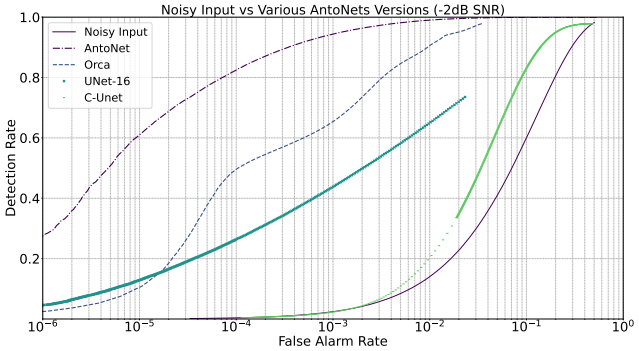

Fig. 7. ROC curves comparing the noisy input with various models output using a pixel-wise detector performance.

V. Conclusion

This paper proposes a lightweight CNN suitable for cheap hardware. The use of linear operations and reduced complexity makes possible the interpretation of the results. Its performance is comparable to bigger models for low SNR.

References

[1] S. Antonio, F. Lionel, and D.-P. Laurent, "Antonet: A small linear contrast-enhancing cnn for pixel-wise detection," in *2024 IEEE Thirteenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2024, pp. 1–6.

[2] D. Pillon, M. Solal, and S. Brasseur, "Simultaneous detection and target motion analysis from conventional passive beamforming outputs," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*. Los Alamitos, CA, USA: IEEE Computer Society, apr 1991, pp. 1321–1324. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/ICASSP.1991.150653

[3] S. Sitbon and P. Blanc-Benon, "Intégration dynamique en sonar passif: détection et localisation simultannées de sources faibles," *TS. Traitement du signal*, vol. 13, no. 5, pp. 449–458, 1996.

[4] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1806–1819, 2011.

[5] W. Tian, Z. Chen, J. Shen, F. Huang, and L. Xu, "Underwater sonar image denoising through nonconvex total variation regularization and generalized kullback–leibler fidelity," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 11, pp. 5237–5251, Nov 2022.

[6] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[7] S. G. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $l^2(r)$," *Transactions of the American Mathematical Society*, vol. 315, pp. 69–87, 1989. [Online]. Available: https://doi.org/10.1090/s0002-9947-1989-1008470-5

[8] K. Nawres, H. Kamel, and E. Noureddine, "Image denoising using wavelets: A powerful tool to overcome some limitations in nuclear imaging," in *2006 2nd International Conference on Information & Communication Technologies*, vol. 1, 2006, pp. 1114–1118.

[9] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[10] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722–732, 2008.

[11] Y. Zahedi, H. Ghafghazi, S. H. Syed Ariffin, and N. M. Kassim, "Feasibility of electromagnetic communication in underwater wireless sensor networks," in *Proceedings of the Underwater Communications Conference*, Nov 2011, full-text available on ResearchGate.

[12] M. A. Ainslie, *Principles of Sonar Performance Modeling*. Springer, 2010.

[13] T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," *J. Acoust. Soc. Am.*, vol. 104, no. 3, pp. 380–387, 1998.

[14] C. Chen and S. Holm, "Acoustic scattering from rough surfaces: A hybrid approach," *Journal of the Acoustical Society of America*, vol. 111, no. 3, pp. 1234–1242, 2002.

[15] L. Dong, L. He, M. Mao, G. Kong, X. Wu, Q. Zhang, X. Cao, and E. Izquierdo, "Cunet: A compact unsupervised network for image classification," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2012–2021, 2018.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[17] C. Bergler, M. Schmitt, A. Maier, S. Smeele, V. Barth, and E. Nöth, "Orca-clean: A deep denoising toolkit for killer whale communication." in *INTERSPEECH*, 2020, pp. 1136–1140.

[18] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[20] R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, "Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 3954–3962, 2018.