DEEP LEARNING BASED TECHNIQUE FOR ENHANCED SONAR IMAGING

Louise Rixon Fuchs[a,b], Christer Larsson[b,c], Andreas Gällström[b,c]

[a]KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, Teknikringen 14, SE-100 44 Stockholm, Sweden
[b]Saab, SE-581 88 Linköping, Sweden
[c] Department of Electrical and Information Technology, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden

Contact author: Louise Rixon Fuchs, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, Teknikringen 14, SE-100 44 Stockholm, Sweden.
Email: rixon@kth.se

Abstract: *Several beamforming techniques can be used to enhance the resolution of sonar images. Beamforming techniques can be divided into two types: data independent beamforming such as the delay-sum-beamformer, and data-dependent methods known as adaptive beamformers. Adaptive beamformers can often achieve higher resolution, but are more sensitive to errors. Several signals are processed from several consecutive pings. The signals are added coherently to achieve the same effect as having a longer array in synthetic aperture sonar (SAS). In general it can be said that a longer array gives a higher image resolution. SAS processing typically requires high navigation accuracy, and physical array-overlap between pings. This restriction on displacement between pings limits the area coverage rate for the vehicle carrying the SAS. We investigate the possibility to enhance sonar images from one ping measurements in this paper. This is done by using state-of-the art techniques from Image-to-Image translation, namely the conditional generative adversarial network (cGAN) Pix2Pix. The cGAN learns a mapping from an input to output image as well as a loss function to train the mapping. We test our concept by training a cGAN on simulated data, going from a short array (low resolution) to a longer array (high resolution). The method is evaluated using measured SAS-data collected by Saab with the experimental platform Sapphires in freshwater Lake Vättern.*

Keywords: *Sonar Imaging, Synthetic Aperture Sonar, Generative Adversarial Networks, Image Enhancement*

INTRODUCTION AND BACKGROUND

Synthetic aperture sonar (SAS) systems have been around since the 80s [1]. They are mostly used for creating high resolution maps of the seafloor, and for finding small objects on it such as in mine countermeasures (MCM) applications. The heavy attenuation of sound in water makes it desirable to have as high signal-to-noise ratio as possible. A high resolution and a high SNR can be achieved by coherently adding measurements from several pings as the sonar platform moves. The sonar image resolution is proportional to the array dimensions, where a longer array gives a higher resolution, in general. For SAS a longer synthetic array is created in order to create high resolution images from a short array sitting on an underwater vehicle. This demands high navigation accuracy since the position of the element in the synthetic array has to be determined with sub-wavelength accuracy. In addition to inertial navigation sensors, signal processing methods are used to correct for deviations from the nominal position. The use of signal processing methods to correct for position errors is commonly referred to as micro-navigation or autofocus. To apply micro-navigation there is often a restriction on the vehicle's displacement between pings, since some techniques require the sub-arrays to overlap between pings.

Numerous enhancement methods for imaging are available. Some examples of well-known methods developed for natural image processing include different filter methods for noise reduction or contrast enhancement [2]. Most work for SAS image enhancement is focused on the image reconstruction using methods for micro-navigation and autofocus [3]. Methods from deep learning have started to be applied for image enhancement purposes in recent years. For example convolutional neural networks (CNNs) have been applied to both natural images [4], medical images [5], RADAR [6], and to SONAR [7]. Generative Adversarial Networks (GANs) have also been used for denoising and image enhancement for medical imaging [5]. Image-to-image translation using GANs has been proposed for RADAR [8], where it was used to transform low resolution SAR imaging to its higher resolution counterpart. Enhancement of one ping SAS images seems to be a new field of research without many publications. In [9] one ping SAS image enhancement is performed using a compressive sensing method.

Our research questions for this paper are: Is it possible to transform a low resolution sonar image to its high resolution counterpart, using state-of-the art methods for Image-to-Image translation, and can sonar image enhancement benefit in other ways from using GANs?

METHODS

The work in this paper can be separated into two parts: one part using simulated data, and another part using measured sonar data. In common for the two parts however is the use of a conditional GAN (Pix2Pix) for paired image-to-image translation tasks. The Pix2Pix network is trained on sonar data generated using simulations.

Generative Adversarial Networks and conditional Generative Adversarial Networks: Generative Adversarial Networks (GANs), introduced by [10] is a framework for estimating generative models. It builds on the idea of having two adversary networks: a generator network G, and a discriminator network D. The generator produces new samples in a distribution in data space $G(z, \theta_g)$, where z is input noise, and θ_g are the parameters of a multilayer perceptron. The discriminator's role is to estimate the probability estimate of the data having been generated as opposed to it being real. Mathematically this process can be defined as having D and G play a two-player min-max game, seen in the equation below,

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))] \quad (1)$$

where $p_z(z)$ is the prior on the input noise variables, $D(x)$ is the probability that the data x came from the data distribution rather than the generator's distribution p_g.

The min-max game ends when D and G cannot improve anymore, and the discriminator cannot distinguish between real, and fake samples, this happens when $p_g = p_{data}$, and $D(x) = 0.5$. There have been many variants developed since the original introduction of GANs in 2014. One of these variants is the conditional GAN (cGAN) which introduces a conditioning y on both G and D [11]. With this extra piece of information it becomes possible to direct the data generation process. The conditioning y could be for example the label of the data.

In this work a conditional GAN [12] is used to generate "artificial" high resolution sonar images from its low-resolution counterpart. This can also be explained as going from a short to a long phased array sensor. We then compare and evaluate the generated "artificial" images to their corresponding ground truth high resolution images. We use the implementation of Pix2Pix provided in [13].

Simulated data: Simulated SAS data is generated at the signal level for the proof of concept. An arbitrary number of point targets of different target strengths can be placed in a scene in the simulator. Parameters that can be defined in the simulator are: sound velocity, array dimensions, number of sonar elements in the array, source level of the transmitter, frequency content and pulse waveform. As a default we use a chirp with the same centre frequency as used in the experimental data acquired by the Sapphires SAS sensor. A SAS image is formed by a using a conventional back-projection algorithm on the individual sonar element signals.

Experiments: We create two kinds of datasets (A and B) for our experiments, where each scene in an image A_i corresponds to a scene in an image B_i. Dataset A holds sonar images created from a short array of 50 elements. The dataset B holds the corresponding sonar images for a longer array with 100 or more elements. The spacing between the elements is always the same so a 100 element long array is twice as long as a 50 element array. When generating the point targets in the scene their location and target strength are updated randomly for each image pair A_i and B_i. We generate 500 images for both A and B, and separate them into training (300 images), validation (100 images), and testing (100 images). The cGAN is then trained to learn how to go from dataset A to B.

Measured SAS data: The used experimental SAS data has been collected with the experimental platform Sapphires in freshwater Lake Vättern. The shorter array image for the measured SAS-data is translated through a cGAN network which has been trained using two datasets A and B (as above) containing simulated point scatters. We use a test image of a tripod processed from measured SAS data (Fig. 3) in our experiments. More details are given in the Results section of the paper.

RESULTS

The experiments performed can be divided into two parts. For the first part simulated point scatter data is used. For the second part, measured SAS data is used and transformed by a cGAN trained on the simulated data from the first experiment.

Simulated data: We study the effect of an Image-to-Image translation task for simulated point targets. The task is to go from a dataset A which holds images formed by a short array to a corresponding dataset B for images formed by a long array. The experiment is performed with one, three and 10 point targets in the respective runs. In each run the cGAN is trained with a dataset that has the same number of point targets in each image. The network is later evaluated on its validation data. An assessment of how well the output image from the cGAN corresponds to its target ground truth image is done using the metrics structural similarity (SSIM) index [14], and mean square error (MSE). The results are shown in Table 1. The SSIM metric is also taken between the datasets A and B as a benchmark to make sure that the high SSIM scores are not due to the datasets A and B being very similar. Note that increasing the number of point targets leads to a degradation in performance between the cGAN output and the ground truth in dataset B, see the reduced SSIM score and increased MSE score (Table 1). The dataset A corresponds to images from a short array 50 elements, and the dataset B corresponds to images from a long array of 100 elements, for this assessment. A visual example for a translation can be seen in Fig. 1 where the short array contains 50 elements and the long array 400 elements.

We also carry out experiments to investigate separation of closely spaced point targets. The cGAN appears to be able to separate closely spaced targets when the two targets have similar target strength (Fig. 2). In the case of having two closely spaced point targets with different target strength, the cGAN frequently translate the two targets as one target point located at the position of the strongest point target in the high-resolution domain. For this experiment the short array has 50 elements, while the long array has 400 elements.

We observe that translating simulated SAS images with more point targets (up to 10) with a cGAN trained on less point targets in each image also works well.

Table 1: Error measures for simulated point targets

SSIM between output image from GAN and corresponding ground truth image from dataset B.				
# Point targets	**SSIM mean**	**SSIM std**	**MSE mean**	**MSE std**
1	0.9862	0.0012	4.3500	4.3500
3	0.9475	0.0166	26.3825	13.2465
10	0.9455	0.0053	37.1433	5.4232
SSIM between A and B dataset for benchmark.				
# Point targets	**SSIM mean**	**SSIM std**	**MSE mean**	**MSE std**
1	0.6428	0.0076	113.4660	4.2450
3	0.7649	0.0324	117.5165	21.2098
10	0.8182	0.0207	125.7439	9.9929

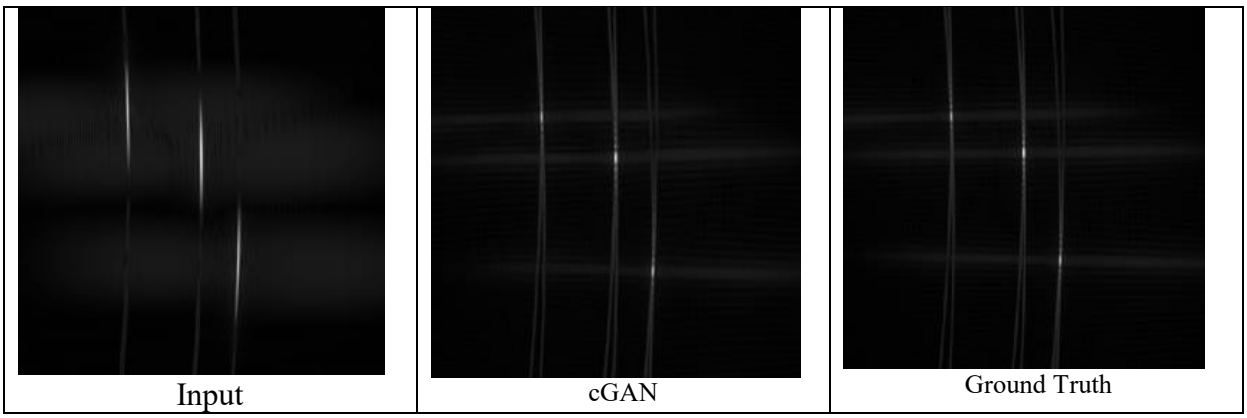

Fig. 1: Point targets. To the left is the depiction from the point targets using the short array. The middle image shows the output from the cGAN. To the right is the ground truth output from the long array.

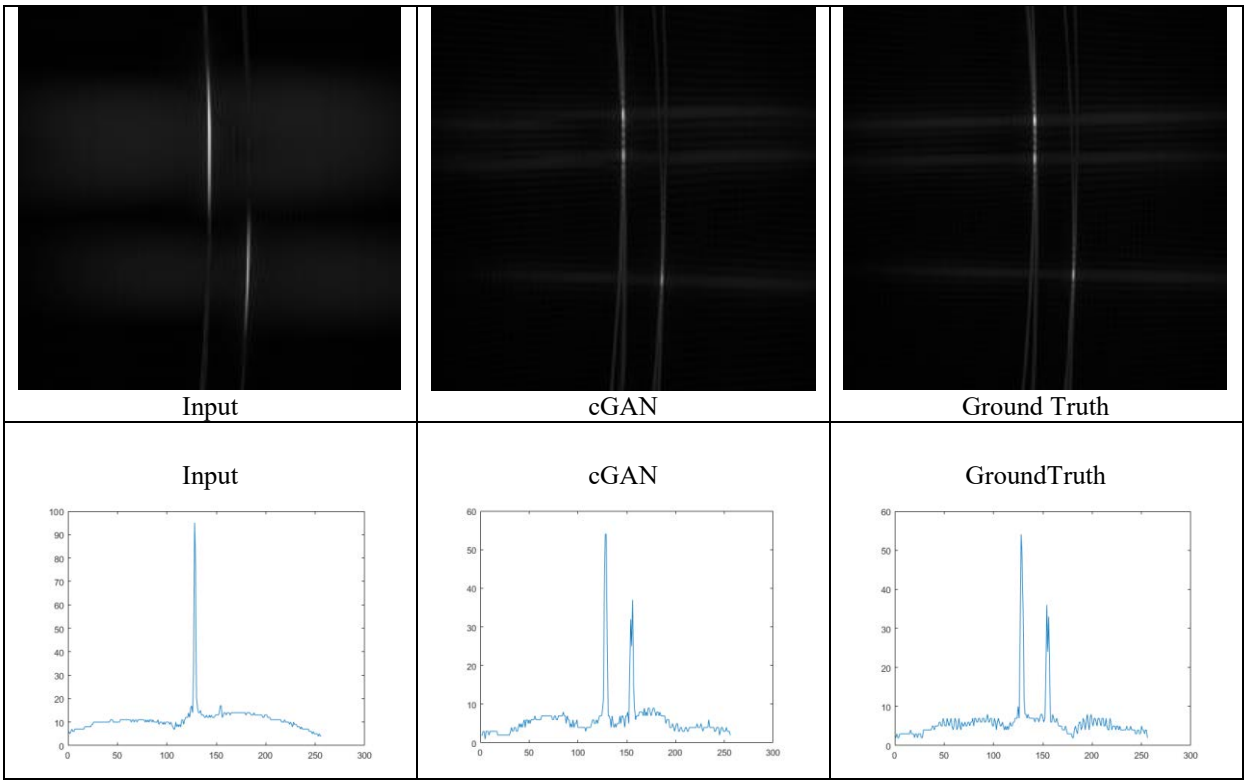

Fig. 2: Separation of closely spaced point targets. The two point targets in the leftmost image are closely spaced and cannot be distinguished as two separate targets. After being sent through the cGAN the output results in an image where the closely spaced targets are distinguishable as separate targets. Upper left (input short array).Upper middle, the output from the cGAN. Upper right (ground truth long array image). The lower row shows the distribution of intensities of a vertical slice from the position of the two closely spaced targets.

Measured Data: The purpose of the second experiment is to investigate if there are any advantages of using the cGAN applied to measured SAS data. We use our test image for all the runs of this experiment. A cGAN trained on scenes containing three point scatters is used. The translated ground truth SAS image appears to have a reduced background level (Fig. 3). For comparison we apply a traditional method for image enhancement, namely that of contrast stretching [2]. In addition we also apply some methods for denoising the image using

a convolutional neural network, a median filter and wavelet-transforms. One the evaluation measures used is entropy as introduced by Shannon [15]:

$$H(x) = -\sum_{i=1}^{N} p_i log_2 p_i, \quad (2)$$

where the p_i values are the probabilities of the different intensity levels for the grey-scaled image. Having a higher entropy number means there is more information in the image. Using entropy minimization can be used for autofocus of ISAR images [16]. Inspired by this we hypothesize that for a target to appear clearly in the image we want to have lower entropy, which should be achieved if signal leakage is removed and the background signal level is suppressed. A ratio between the background and target (Target to background ratio –TBR) is also calculated through taking the ratio between the mean value of the pixel intensities for an area containing the target and the mean pixel intensities for an area containing background. The same areas for background and target are used for all images in order to achieve comparable measurements. The results are summarized in Table 2. We can conclude that the use of cGAN reduces the entropy for the image, and gives the highest target-background ratio. The entropy was lowered further through pre-processing using Otsu thresholding [17] (with four threshold values).

Additionally we analyse the cGAN effect on the input image using the SSIM metric (Table 3). Note that the cGAN's SSIM score between the cGAN processed 50-element image, and the 100-element is higher than the SSIM score between the original 50-element image and the 100-element image. For this experiment the GAN was trained to go from 50-elements to 100-elements for scenes with three point targets.

Table 2 Evaluation measures for our SAS-test image made from 50 elements (leftmost column). The following columns shows evaluation measures for the SAS-test image processed in different ways, the rightmost columns holds the evaluation measures for a SAS-test image made from 100 elements.

	50-element	cGAN	Thres	Thres+cGAN	Median	CNN	Contrast	Wavelet	100-elements
Entropy	6.6957	6.0921	**5.0124**	5.3716	5.9507	6.0827	6.0169	6.0930	5.6615
TBR	1.5079	2.6287	4.1349	**4.8119**	1.5474	1.5087	2.1102	1.5080	1.6988

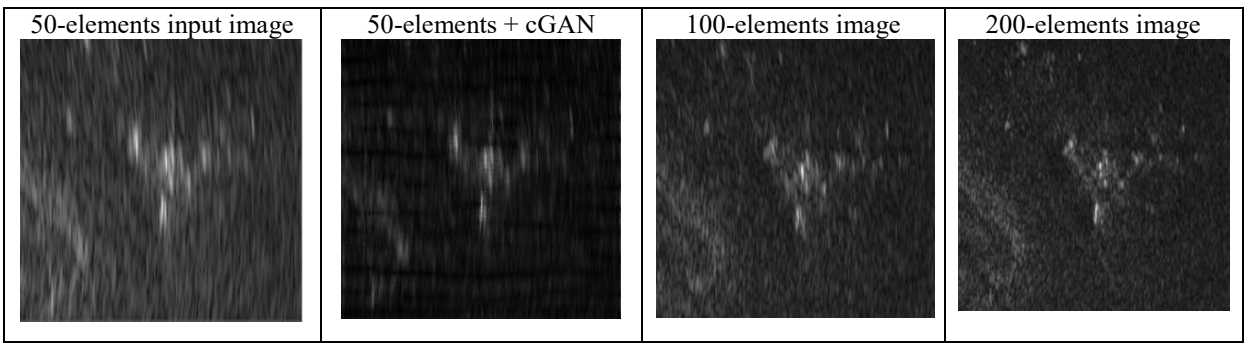

Figure 3 Input SAS image from 50-elements (left), output image after the cGAN (middle-left). SAS image from 100-elements (middle-right), and SAS image from 200-elements (right).

Table 3 SSIM for comparison of measured SAS data and cGAN

	50-element image	cGAN	100-element image
100-element image	0.3197	0.3456	1
200-element image	0.3205	0.3514	0.6232

CONCLUSION

We show that cGAN is a promising tool to enhance sonar imaging. In particular our contributions in this work are the following:

1. We demonstrate some capability for separation of point targets with cGANs.
2. We show that it is feasible to go from a low resolution sonar image produced from a (short-array) to its high resolution (long-array) counterpart with simulated data.

It is hard to conclude at this point that this method could lead to a relaxed requirement for array overlap for SAS in order to enable a higher platform speed. Some increased detection capabilities for sonar targets using cGANs for enhancement may also be possible. For further work it could be interesting to train the whole cGAN entirely on measured SAS data going from a short to long array.

ACKNOWLEDGEMENTS

This project is financed by the Swedish Maritime Robotic Centre (SMaRC) (IRC15-0046), a national centre funded by the Swedish Foundation for Strategic Research (SSF). This work was also supported by SSF through Strategic Mobility grants.

REFERENCES

[1] **M. P. Hayes and P. T. Gough,** Synthetic aperture sonar: A review of current status, *IEEE Journal of Oceanic Engineering*, 34(3), pp. 207-224, July 2009.
[2] **R. C. Gonzalez and R. E. Woods**, *Digital Image Processing (3rd Edition),* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
[3] **H. J. Callow.** *Signal processing for synthetic aperture sonar image enhancement*, PhD thesis, University of Canterbury, 2003.
[4] **V. Jain and S. Seung**, Natural image denoising with convolutional Networks, In *D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, Advances in Neural Information Processing Systems 21,* pages 769-776. Curran Associates, Inc., 2009.
[5] **Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K Kalra, Y. Zhang, L. Sun, and G. Wang,** Low-dose ct image denoising using a generative adversarial network with

wasserstein distance and perceptual loss, *IEEE Transactions on Medical Imaging*, 37(6), pp 1348-1357, 2018.

[6] **P. Wang, H. Zhang, and V. M Patel**, Sar image despeckling using a convolutional neural network, *IEEE Signal Processing Letters*, 24(12), pp. 1763-1767, 2017

[7] **M. Sung, H. Joe, J. Kim, and S. Yu,** Convolutional neural network based resolution enhancement of underwater sonar image without losing working range of sonar sensors, In *2018 OCEANS -MTS/IEEE Kobe Techno-Oceans (OTO)*, pp.1-6, 2018.

[8] **D. Ao, CO Dumitru, G. Schwarz, and M. Datcu**, Dialectical gan for sar image translation: From sentinel-1 to terrasar-x, *Remote Sensing*, 10(10), 2018.

[9] **A. Gällström, L. Rixon Fuchs and C. Larsson**, Enhanced sonar image resolution using compressive sensing modelling, In *UACE 2019, Crete-Greece*, 2019.

[10] **I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D.Warde-Farley, S. Ozair, A. Courville, and Y. Bengio**, Generative adversarial nets, In *NIPS*, 2014.

[11] **M. Mirza and S. Osindero,** Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784***,** 2014.

[12] **P. Isola, J. Zhu, T. Zhou, A. A. Efros,** Image-to-image translation with conditional adversarial networks**,** In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *Honolulu, HI, USA, July 21-16, 2017*, pp. 5967-5976, 2017.

[13] Implementation of pix2pix. https://github.com/affinelayer/pix2pix-tensorflow. Accessed: 2019-03-26.

[14] **Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli**, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 600-612, Apr. 2004.

[15] **Shannon, C. E.** , A mathematical theory of communication, *Bell Systems Technical Journal,* 27:379-423, 623-656, 1948.

[16] **L. Xi, L. Guosui and J. Ni**, Autofocusing of ISAR Images Based on Entropy Minimization, *IEEE Transactions on Aerospace and Electronic Systems,* vol. 35, no. 4, October 1999.

[17] **Otsu, N.**, A Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems, Man, and Cybernetics*. Vol. 9, No. 1, 1979, pp. 62–66.