Generalization of a deep learning network for beamforming and segmentation of ultrasound images

Generalization of a deep learning network for beamforming and segmentation of ultrasound images

Silvia Seoni
Polito^{BIO}Med Lab, Biolab, Dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
silvia.seoni@polito.it

Giulia Matrone
Dept. of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
giulia.matrone@unipv.it

Nicola Casali
Polito^{BIO}Med Lab, Biolab, Dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
nicola.casali@polito.it

Edoardo Spairani
Dept. of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
edoardo.spairani01@universitadipavia.it

Kristen M. Meiburger
Polito^{BIO}Med Lab, Biolab, Dept. of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
kristen.meiburger@polito.it

***Abstract*— Recently, deep neural networks (DNNs) for beamforming and segmenting plane-wave ultrasound images have been proposed. The promising results obtained so far focus on segmenting anechoic, almost circular structures using one architecture trained on a large dataset. We present a study of DNNs generalizability for beamforming and segmenting structures of various shapes and echogenicity. Three different encoder architectures (i.e. VGG13/16/19) and target images with standard dynamic range (dR = 60 dB, E_{60}) or an automatically determined dR (E_{auto}) were compared. Field II was used to simulate 6560 images (with hyperechoic, hypoechoic, anechoic and mixed targets) using random bunches of ellipses to generate different shapes for DNN training. The test set included 816 simulated images, 21 images of a phantom (CIRS040GSE) and 24 images of the carotid artery. The DNN architecture has 1 encoder and 2 decoders, for segmentation and beamforming, based on the UNet. Using the VGG19 trained with E_{auto} images, a considerable improvement was achieved when compared to other architectures, especially when performing tests on experimental data. Overall, the promising results obtained encourage us to further investigate the use of DNNs for beamforming and segmentation, with the aim to improve the performance and generalize their use for specific ultrasound imaging applications.**

Keywords—Deep learning, Beamforming, Segmentation, Dynamic range, Deep neural networks

I. Introduction

Ultrasound images are widely used to support diagnosis in the clinical field. Deep learning approaches have skyrocketed over the last years and have been applied to various ultrasound tasks, such as classification, segmentation and image quality assessment [1]. Recently, deep neural networks (DNNs) have also been proposed for beamforming [2],[3],[4],[5],[6] and segmentation of plane-wave ultrasound images [3]. Indeed, starting from the raw data, it is possible to obtain information and directly generate the ultrasound image and a segmentation map [3]. Another innovation is the introduction of the generative adversarial network (GAN) model into the field of ultrasound image reconstruction, as an alternative beamforming method [7], [8]. Recent promising results by Nair et al. [3] segment anechoic, almost circular structures, without training the network to consider structures with different shapes or echogenicity. Furthermore, the network was often trained using a very large dataset, which involves a high computational cost and time. Given the different shapes and echogenicity of anatomical structures, the generalization of DNNs for simultaneous beamforming and segmentation is necessary to optimize their application. Moreover, different beamforming methods may alter the dynamic range (dR) of B-mode images. Our previous study [9] proposed an innovative method to estimate an automatic dR based on the image histogram.

We present here a study of DNNs' generalizability for beamforming and segmenting structures of various shapes and echogenicity. We compare the performances using 3 different encoder architectures and a standard dR (60 dB) or an automatically-determined dR for the target images. The following sections are organized as follows. Section II describes the network architecture, the dataset, and the training modality. Section III reports the results obtained on our dataset. Section IV includes a discussion about the results and future works.

II. Materials and Methods

We aim to produce a DNN beamformed image and a segmentation mask prediction, starting from raw data (phase and quadrature IQ data).

A. Ultrasound images dataset

The dataset was composed of simulated and experimental plane-wave ultrasound images. A total of 8192 simulated images (6560 in the training set, 816 in the validation set and 816 in the test set) were generated using Field II and Matlab 2019b software. The simulation generated the IQ data, the RF data, the beamformed image (using the Delay and Sum (DAS) method), the segmentation map and the enhanced image (the combination of the B-mode image and segmentation mask). The images could contain individual or double random bunches of ellipses. The radius, the axial and lateral position of the simulated cysts were changed to increase the data variability. The radius varied randomly in the range of 2-10 mm. These cysts were contained within a cuboidal phantom volume located between an axial depth of 10 mm and 50 mm, with a lateral width of 32 mm, and an elevational thickness of 1 mm. To have fully developed speckle, 10000 scatterers were contained in the simulated phantom. All images were simulated with an attenuation coefficient of 1.5 dB/(cm MHz). The simulated transducer had 128 elements working at 5 MHz. The structures were simulated by varying the echogenicity of the inclusions (i.e., hyperechoic, hypoechoic, anechoic, and mixed targets).

The experimental ultrasound images included 24 phantom images (model 040GSE, CIRS Inc, USA) and 21 *in vivo* carotid images, acquired on two healthy volunteers. A Verasonics Vantage 128 system (with a L11-5v linear array) was used for experimental plane-wave acquisitions. The phantom contains hyperechoic, hypoechoic and anechoic targets, that are placed at a depth between 5 and 60 mm. Examples of the simulated and experimental test set images are displayed in Fig. 1.

The standard 60 dB value was used as a dynamic range for all images. Furthermore, an automatic dynamic range was also computed [9] for all images. Indeed, the different target positions and echogenicity may alter the image dynamic range and further processing. Thus, to avoid this problem, we used the automatic dR, which may vary between the images, to obtain visually similar results.

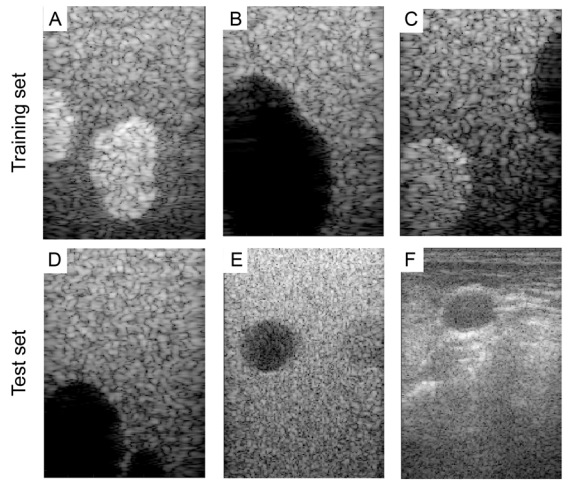

Fig. 1. Top row: examples of the training set images with hyperechoic (A) hypoechoic (B) and mixed echogenicity cysts (C). Bottom row: examples of the test set, including simulated images (D), phantom images (E) and *in vivo* carotid images (F).

B. Deep learning network architecture and training

The DNN architecture proposed in [3] is based on the UNet [10] for biomedical image segmentation, which consists of a contraction path and two expansive paths: one for the DNN image and the other for the segmentation map. We tested three different encoders, i.e. the VGG13, VGG16 and VGG19 architectures, which differ in the number of weight layers (13, 16 and 19 respectively) [11]. The DNNs were trained for 50 epochs, with a learning rate of $5*10^{-5}$ and using Adam optimizer. The minibatch size was set to 8. An early-stopping criterion was set, to stop the training if the validation loss did not change for 6 consecutive epochs. For data augmentation, each plane-wave acquisition was flipped from left to right. The input data were rescaled to ensure that the values lie in a range of 0-1.

The networks were trained using four different targets: the enhanced beamformed image with a dynamic range of 60 dB (E_{60}), the enhanced image with an automatically determined dR (E_{auto}), the original beamformed image with a dynamic range of 60 dB (B-mode$_{60}$) or again with the automatically determined dR (B-mode$_{auto}$). For the enhanced images, we considered the enhancement of the original image with a segmentation mask only for the anechoic cyst cases. Instead, when considering hyperechoic or hypoechoic cysts, we did not enhance the target image and used the original beamformed one.

The networks were trained to learn the reference, that is composed of the combination of the segmentation map and the beamformed image (i.e., E_{60}/E_{auto}/B-mode$_{60}$/B-mode$_{auto}$) obtained from the input signals; it then generates the DNN image and DNN segmentation mask predictions. The loss function was defined as the sum of the following two losses: the mean absolute error and the Dice similarity coefficient (DSC) [3]. The first loss was computed as the mean absolute error between the predicted DNN image D and the reference beamformed image. As mentioned before, the latter could be either the enhanced beamformed image E_{60}/E_{auto} (in the case of anechoic targets) or the original beamformed image B-mode$_{60}$/B-mode$_{auto}$ (for the hypo/hyperechoic targets). The second loss was the Dice Similarity coefficient between the predicted DNN segmentation and the true segmentation.

For testing the DNNs, 816 simulated images, 24 images of the phantom, and 21 images of the carotid artery were used. Segmentation performance was analysed using the DSC, while the DNN beamformed image was evaluated using contrast (Con), signal to noise ratio (SNR), generalized contrast to noise ratio (gCNR) and peak SNR (PSNR) [3]. The DSC [12] quantifies the overlap between the predicted and real segmentation mask. To compute the contrast, two ROIs (circles with radius of 2 mm) were placed on the image as follows: a first ROI was placed inside the cyst and the second outside, at the same depth of the cyst center. The signal-to-noise ratio (SNR) [13] quantifies the smoothness of the background region near the cyst. The generalized contrast to noise ratio (gCNR) [14] is an accurate index of lesion detectability. Furthermore, the peak SNR (PNSR) [13], [15] was computed to quantify the similarity between the DNN image and the beamformed image. The validation metrics were estimated for each network (VGG13,

TABLE I. VALIDATION METRICS IN THE SIMULATED CYST IMAGES (VGG19 TRAINED WITH E_{AUTO})

Metrics	Simulated hypo/anechoic		Simulated hyperechoic	
	DAS	***DNN***	***DAS***	***DNN***
DSC	/	0.96±0.09	/	0.97±0.02
Con (dB)	-12.67±9.08	-54.64±40.6	6.05±2.59	6.00±2.96
SNR	4.24±1.67	5.48±1.67	3.79±1.70	7.32±3.13
gCNR	0.58±0.20	0.63±0.23	0.37±0.10	0.39±0.10
PSNR (dB)	/	24.33±0.94	/	23.78±1.21

VGG16 and VGG19), using the four different targets(E_{60}/E_{auto}/B-mode$_{60}$/B-mode$_{auto}$).

III. RESULTS

The validation metrics are listed in Table I for the simulated images and in Tables II and III for experimental datasets. For the simulated and phantom images, the metrics were reported for the hypo/anechoic cyst and the hyperechoic cyst separately. Fig. 2 shows an example of DAS and DNN images and the two segmentation masks for two different simulated images, one phantom image, and a carotid image.

With the simulated image test set, the validation parameters were comparable between the three architectures. In these images with hypo/anechoic cysts, the test set produced mean ± standard deviation DSC of 0.96±0.09, contrast of -54.64±2.96 dB, SNR of 5.48±1.67, and PSNR of 24.33±0.94 dB using the VGG19 trained with E_{auto}. For the hyperechoic cyst, we obtained DSC=0.97±0.02, Con=6.00±2.96 dB, SNR=7.32±3.13, gCNR=0.39±0.10, PSNR 23.79±1.21 dB for the hyperechoic cyst (again using VGG19 with E_{auto}). The results obtained using the other two encoders (i.e., VGG13 and VGG16) were similar when considering the simulated dataset. Hence, Table I reports only the metrics obtained with the network VGG19 trained with E_{auto} target images.

TABLE II. VALIDATION METRICS IN THE PHANTOM IMAGES

	Metrics	VGG13		VGG16		VGG19	
		DAS	***DNN***	***DAS***	***DNN***	***DAS***	***DNN***
Hypo/anechoic cysts	DSC	/	0.11 ± 0.07	/	0.10± 0.15	/	**0.54 ± 0.18**
	Con (dB)	-5.42± 3.87	-2.50± 1.90	-5.42± 3.87	-8.13± 8.53	-5.43± 3.87	**-31.7± 33.03**
	SNR	5.07± 0.54	7.86± 0.87	5.07± 0.54	**9.68± 1.96**	5.07± 0.54	9.40± 0.79
	gCNR	0.38± 0.09	0.33± 0.03	0.38± 0.09	0.42± 0.14	0.38± 0.09	**0.57± 0.26**
	PSNR (dB)	/	16.97± 0.33	/	16.83± 0.48	/	**17.17± 0.80**
Hyperechoic cysts	DSC	/	0.25 ± 0.02	/	0.29± 0.26	/	**0.76± 0.14**
	Con (dB)	2.22± 0.49	1.36± 0.30	2.22± 0.49	**1.64± 0.12**	2.22± 0.49	1.55± 0.24
	SNR	5.65± 0.47	9.54± 0.94	5.65± 0.47	**11.49± 1.27**	5.65± 0.47	10.74± 1.88
	gCNR	0.34± 0.06	0.29± 0.03	0.34± 0.06	0.28± 0.04	0.34± 0.06	**0.37± 0.07**
	PSNR (dB)	/	17.91± 0.23	/	17.84± 0.24	/	**18.16± 0.43**

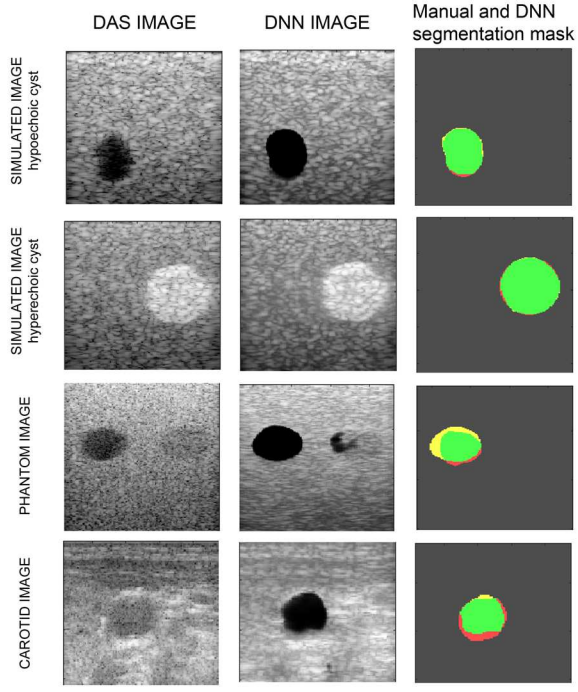

Fig. 2. Comparison between the DAS and DNN images: simulated images (first and second row), real phantom images (third row) and carotid images (last row). The area segmented both manually and by DNN is represented in green; the undersegmented and oversegmented areas by the DNN as compared to the manual mask are in red and yellow, respectively.

For the experimental phantom case, the VGG19 trained with E_{auto} images showed the best results. Tables II shows the validation metrics of the VGG13, VGG16 and VGG19 trained with E_{auto}, for the hypo/anechoic cysts and hyperechoic cysts, respectively. A considerable improvement with the VGG19 encoder can be observed, especially for the DSC value. Indeed, the segmentation considerably improved both for the hypoechoic (DSC=0.54±0.18 vs 0.10±0.15 and 0.11±0.07 with VGG19, VGG16 and VGG13, respectively) and hyperechoic cysts (DSC=0.76±0.18 vs 0.29±0.14 and 0.25±0.02 with VGG19, VGG16 and VGG13, respectively).

For the *in vivo* carotid images, the VGG19 architecture trained with automatic-dR images provided a performance increase for the DSC, contrast and PSNR parameters, when compared to VGG13 (i.e., DSC=0.72±0.22 vs 0.01±0.01, Con=-5.47±2.84 dB vs -3.68±9.00 dB, SNR=6.29±1.12 vs 8.28±1.08, gCNR=0.34±0.10 vs 0.41±0.13, PSNR=19.01±0.67 dB vs 15.40±0.77 dB). Due to the variability in echogenicity in the carotid images, the DNN often segmented other areas of interest. To mitigate this, post-processing was done on the segmentation mask, based on determining the most hypoechoic areas on the DNN images within the obtained segmentation mask and then performing region growing. This allowed a segmentation performance (DSC) increase from 0.08±0.09 to 0.72±0.22 (with VGG19).

In general, the performance of the DNNs trained with different targets did not show significant differences. Indeed, the

TABLE III. VALIDATION METRICS IN THE CAROTID IMAGES (VGG19 TRAINED WITH E_{AUTO})

	Metrics	*DAS*	*DNN*
Carotid images	DSC	/	0.72±0.22
	Con (dB)	-4.43±1.65	-5.47±2.84
	SNR	4.18±0.55	6.29±1.12
	gCNR	0.32±0.06	0.34±0.10
	PSNR (dB)	/	19.01±0.67

training using *E* only resulted in an improvement of the contrast in the experimental phantom images with hypo/anechoic cysts (Con=-14.66±15.78 dB, -31.82±33.03 dB, -3.96±2.35 dB and -3.49±2.25 dB with E_{60}, E_{auto}, B-mode$_{60}$ and B-mode$_{auto}$, respectively) and in the simulated images with hypo/anechoic cysts (Con=-55.17±40.44 dB, -54.64±40.64 dB, -12.77±9.85 dB and -14.31±10.88 dB with E_{60}, E_{auto}, B-mode$_{60}$ and B-mode$_{auto}$, respectively), using VGG19.

IV. Discussion

The results presented in this study show the feasibility of training DNNs for the beamforming and segmentation of ultrasound images that contain structures with different echogenicity and shapes. In addition to demonstrating the ability of DNNs to distinguish the background from cysts, we also demonstrated their ability to distinguish cysts with different echogenicity present in the same image. Testing different encoders, we were able to improve the results obtained with VGG13 by using the VGG19 architecture, especially in the case of *in vivo* carotid images. For simulations, the DNN with VGG19 trained with E_{auto} images obtained similar performances as when using the VGG13 and VGG16 encoders. However, the major improvement using the VGG19 was for the experimental images acquired on a phantom and on healthy volunteers. Indeed, the network manages to generate better segmentation maps and beamformed images with higher contrast and SNR. Furthermore, the novel use of the automatic dR [9] for the target images used in the training phase shows a small improvement of the performance, especially for the contrast and DSC in the phantom images with hypo/anechoic cysts. Another key observation should be made for the dimension of the training set. Indeed, the results were obtained using a much smaller training set compared to [3] (6560 here vs. 22230), which reduces the computational time and cost.

Given the encouraging results obtained, future work could investigate the use of a dataset with an even higher variability of structures, as well as the inclusion of experimental images in the training dataset. Indeed, in this work the DNNs were able to generate images with specific structures; the use of experimental images in the training set could be a key factor to improve performance further and to generalize the experimental applications of DNNs. Furthermore, considering the influence of the encoder architecture on the obtained results, future works could focus on the training of the network with different architectures, both for the encoder and the decoder.

V. Conclusion

This work shows that a DNN can be used both for beamforming and segmenting ultrasound images with structures having different shape and echogenicity. Using VGG19 as an encoder, we were also able to improve the performance on experimental data. Overall, promising results have been obtained, which encourage analyses aimed at further generalizing the use of DNNs for ultrasound image reconstruction and segmentation.

References

[1] S. Liu et al., "Deep Learning in Medical Ultrasound Analysis: A Review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019, doi: 10.1016/j.eng.2018.11.020.

[2] S. Khan, J. Huh, and J. C. Ye, "Adaptive and Compressive Beamforming Using Deep Learning for Medical Ultrasound," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, no. 8, pp. 1558–1572, 2020, doi: 10.1109/TUFFC.2020.2977202..

[3] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. Lediju Bell, "Deep Learning to Obtain Simultaneous Image and Segmentation Outputs from a Single Input of Raw Ultrasound Channel Data," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, issue 12, pp 2493-2509, doi: 10.1109/TUFFC.2020.2993779

[4] A. C. Luchies and B. C. Byram, "Deep Neural Networks for Ultrasound Beamforming," *IEEE Trans. Med. Imaging*, vol. 37, no. 9, pp. 2010–2021, 2018, doi: 10.1109/TMI.2018.2809641.

[5] B. Luijten et al., "Adaptive Ultrasound Beamforming Using Deep Learning," *IEEE Trans. Med. Imaging*, vol. 39, no. 12, pp. 3967–3978, 2020, doi: 10.1109/TMI.2020.3008537.

[6] B. Luijten et al., "Deep Learning for Fast Adaptive Beamforming," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 1333–1337, 2019, doi: 10.1109/ICASSP.2019.8683478

[7] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell, "A Generative Adversarial Neural Network for Beamforming Ultrasound Images: Invited Presentation," *2019 53rd Annu. Conf. Inf. Sci. Syst. CISS 2019*, pp. 1–6, 2019, doi: 10.1109/CISS.2019.8692835.

[8] S. Goudarzi, A. Asif, and H. Rivaz, "Multi-focus ultrasound imaging using generative adversarial networks," *Proc. Int. Symp. Biomed. Imaging (ISBI)*, 2019, pp. 1118–1121, doi: 10.1109/ISBI.2019.8759216.

[9] K. M. Meiburger, S. Seoni and G. Matrone, "Automatic Dynamic Range Estimation for Ultrasound Image Visualization and Processing," 2020 IEEE Int. Ultrason. Symp. (IUS), 2020, pp. 1-4, doi: 10.1109/IUS46767.2020.9251470.

[10] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. 3rd Int. Conf. Learn. Represent. ICLR*, 2015, pp. 1–14

[12] K. H. Zou et al., "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004, doi: 10.1016/S1076-6332(03)00671-8.

[13] S. Yao, W. Lin, E. P. Ong, and Z. Lu, "Contrast signal-to-noise ratio for image quality assessment," *Proc. International Conference on Image Processing, ICIP*, vol. 1. pp. 397–400, 2005, doi: 10.1109/ICIP.2005.1529771.

[14] A. Rodriguez-Molares et al., "The Generalized Contrast-to-Noise Ratio: A Formal Definition for Lesion Detectability," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, no. 4, pp. 745–759, 2020, doi: 10.1109/TUFFC.2019.2956855.

[15] A. J. Dinu, R. Ganesan, A. A. Kebede, and B. Veerasamy, "Performance analysis and comparison of medical image compression techniques," *2016 Int. Conf. Control Instrum. Commun. Comput. Technol.* ICCICCT 2016, pp. 738–745, 2017, doi: 10.1109/ICCICCT.20