DEEP PROXIMAL UNFOLDING FOR IMAGE RECOVERY FROM UNDER-SAMPLED CHANNEL DATA IN INTRAVASCULAR ULTRASOUND

Nishith Chennakeshava[⋆], *Tristan S.W. Stevens*[⋆], *Frederik J. de Bruijn*[†], *Andrew Hancock*[†], *Martin Pekař*[†], *Yonina C. Eldar*[‡], *Massimo Mischi*[⋆], *Ruud J.G. van Sloun*[⋆†]

[⋆]Eindhoven University of Technology, [†]Philips Research, [‡]The Weizmann Institute of Science

ABSTRACT

Intravascular UltraSound (IVUS) is a key tool in guiding the treatment and diagnosis of various coronary heart diseases. However, due to its nature IVUS is a very challenging modality to interpret, and suffers from a severely restricted data transfer rate. This forces a trade-off between temporal and spatial resolution. Here, we propose a model-based deep learning solution that aims to reconstruct images from data that has been beamformed by under-sampling the number of channels by a factor of 4. By exploiting the physics based measurement model, we achieve better performance and consistency in our predictions when compared to a benchmark model. This lowers the computational load on existing hardware and enables in exploring our ability to run multiple visualisation modalities simultaneously, without a loss of temporal resolution.

***Index Terms*—** IVUS, AI, Model Based Neural Network, Denoising

1. INTRODUCTION

Intravascular ultrasound (IVUS) has proven to be a valuable tool for accurate diagnosis of diseases and complications that cannot otherwise be imaged by conventional ultrasound. Conditions such as abdominal aortic aneurysm [1], or atherosclerosis [2] employ IVUS as a fundamental tool. Due to IVUS's nature as a cost-effective alternative to many other imaging modalities, it is widely employed, albeit extremely difficult to interpret.

As a catheter-based ultrasound device, IVUS suffers from a highly constricted bandwidth. Therefore, it is forced into a trade-off between temporal and spatial resolution. In order to relax this trade-off, we propose a solution that sub-samples the acquired data by a factor of 4. We then attempt to reconstruct the fully-sampled image through a deep Unfolded Proximal Gradient Network (UPGN), a model-based neural architecture.

Neural networks such as UNets [3], and ResNets [4] are popular choices for tackling similar problems. Such networks and variations of them offer an off-the-shelf solution that works extremely well in a large variety of problems related to medical imaging [5, 6, 7]. Often, these networks are also successfully employed within an Adversarial setting [8].

Unfortunately, these advantages come at a price. These networks often yield solutions that violate the physical measurement model. They are also typically over-parameterised [9], which can lead to problems regarding under-specification [10], which may lead to networks learning "shortcuts" to optimise the problem at hand. We aim to overcome these problems by embedding a model-based approach into the architecture of a neural network, thereby limiting the degrees of freedom in the neural architecture. We achieve this by unrolling a proximal gradient solution, and learning all free parameters from data.

By doing so, we can relax the requirement on pre-existing hardware, allowing for multiple visualisation modalities to be run simultaneously, without losing temporal resolution. This study builds on the work presented in [11] as an inspiration for the training strategy, and utilises it to solve a different problem, specifically that of IVUS imaging.

We begin with deriving the model-based neural architecture in Section 2, followed by an explanation of the data that was used to train the proposed network in Section 3. Next, the training strategy is explained in Section 4. We present the results and discuss them in Section 5. Finally, our conclusions are summarised in Section 6.

2. METHODS

We model the recovery of an IVUS image as an inverse problem in which we aim to reconstruct the underlying fully sampled image from an under-sampled measurement. We express $\mathbf{x} \in \mathbb{R}^N$ as the fully sampled beamformed image, and $\mathbf{y} \in \mathbb{R}^N$ as the beamformed image after under-sampling the number of receive elements. This can be modelled as

$$\mathbf{y} = \mathbf{D}_{sub}\mathbf{S}(\mathbf{x}'), \tag{1}$$

where $\mathbf{D}_{sub}$ is a sub-sampling beamforming matrix, and $\mathbf{S}$ is a non-linear scattering function, acting on the tissue intensity map $\mathbf{x}'$. We then make a linear assumption on the right hand side of (1), giving us,

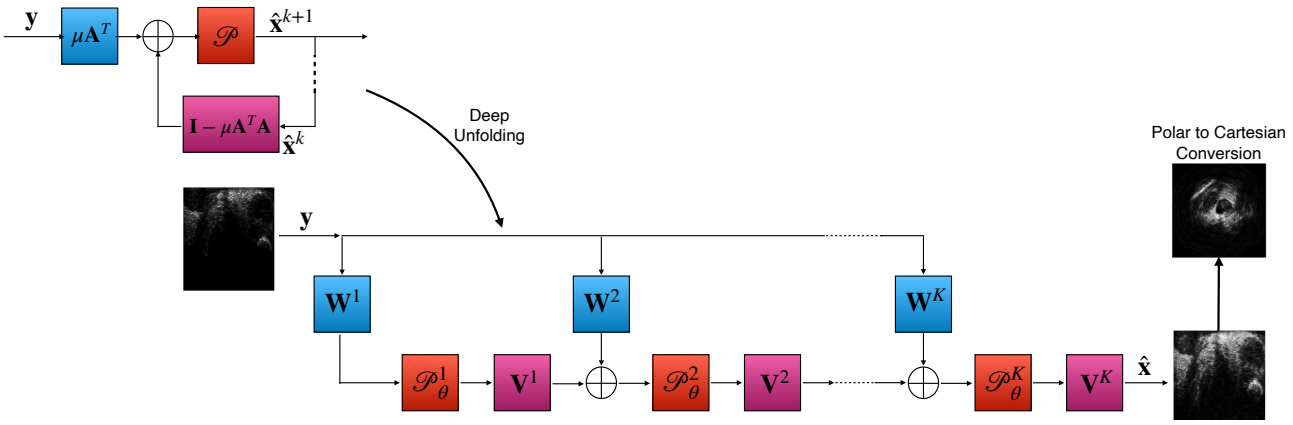

Fig. 1. The image describes the process of unfolding an iterative solution to K folds, giving us a model with fixed computational complexity, and where all free parameters may be learnt from data. We input log-compressed polar data, and then convert the prediction into the cartesian domain for display.

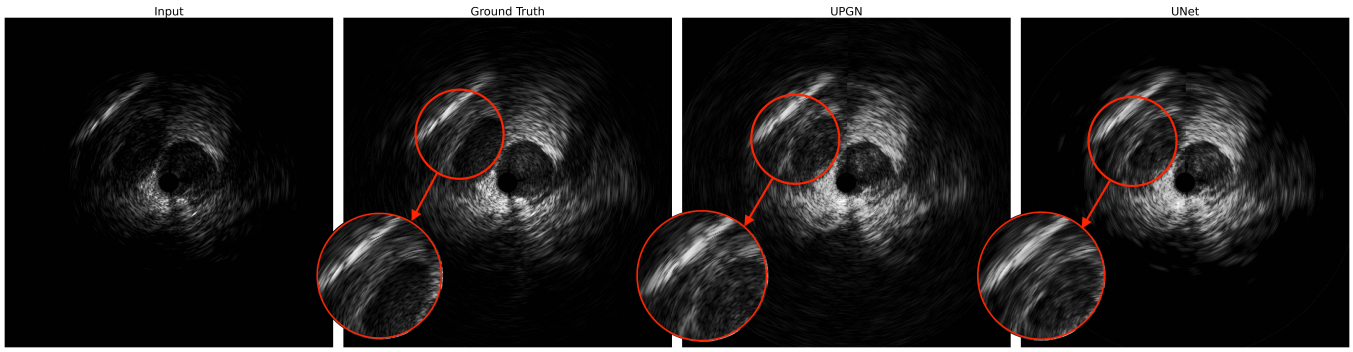

Fig. 2. The images show an under-sampled input, the given ground truth, the proposed Unfolded Proximal Gradient Network (UPGN), and the best of the benchmark networks, the UNet Lite. It also shows an area with the images, that has been focused on.

$$\mathbf{y} \approx \mathbf{A}\mathbf{x}, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the measurement matrix, which may be looked as a degradation matrix in this context. We also assume that the measurement matrix follows a convolutional toeplitz structure, which enables us to re-write $\mathbf{A}$ as a convolutional kernel.

In this context, the measurement matrix can be we written out exactly. However, rather than deriving an analytical solution, we would prefer to learn this from data. Assuming a Gaussian model, we solve this problem by re-writing (2) as,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + R(\mathbf{x}), \tag{3}$$

where $R(\mathbf{x})$ acts as a regulariser.

Assuming that $R(\mathbf{x})$ is known, we may derive a proximal gradient solution for (3). This results in an iterative algorithm that alternates between a data consistency step, and a proximal step that nudges the intermediate solutions to the proximity of the regulariser. Such an algorithm can be described as:

$$\hat{\mathbf{x}}^{(k+1)} = \mathcal{P}(\hat{\mathbf{x}}^{(k)} - \mu\mathbf{A}^T(\mathbf{A}\hat{\mathbf{x}}^{(k)} - \mathbf{y})), \tag{4}$$

where μ is a step size, and $\mathcal{P}^{(k)}$ is the proximal operator of the regulariser [12]. We can now rewrite (4) to separate the contributions of $\hat{\mathbf{x}}$ and $\mathbf{y}$, giving us,

$$\hat{\mathbf{x}}^{(k+1)} = \mathcal{P}_\theta^{(k)}(\mathbf{W}^{(k)}\mathbf{y} + \mathbf{V}^{(k)}\hat{\mathbf{x}}^{(k)}), \tag{5}$$

where θ denotes the learnable parameters of the proximal operator, $\mathbf{W}^{(k)} = \mu^{(k)}\mathbf{A}^T$ and $\mathbf{V}^{(k)} = \mathbf{I} - \mu^{(k)}\mathbf{A}^T\mathbf{A}$. Additionally, by utilising the convolutional nature of $\mathbf{A}$, we may re-write $\mathbf{W}^{(k)}$ and $\mathbf{V}^{(k)}\hat{\mathbf{x}}^{(k)}$ as:

$$\mathbf{W}^{(k)}\mathbf{y} = \mathbf{w}^{(k)} \circledast \mathbf{y}, \tag{6}$$

$$\mathbf{V}^{(k)}\hat{\mathbf{x}}^{(k)} = \mathbf{v}^{(k)} \circledast \hat{\mathbf{x}}^{(k)}, \tag{7}$$

where $\circledast$ represents a convolutional operation, where the convolutional kernels have a size of 3×3. Here, $\mathcal{P}_\theta^{(k)}$ is modelled using a U-Net style NN comprised of 9 convolutional layers

Table 1. A quantitative comparison between the proposed Unfolded Proximal Gradient Network (5 folds), a UNet Lite (a network with a similar number of parameters), a full sized UNet, and a ResNet. We compare the Peak Signal to Noise Ratio (PSNR), the Mean Absolute Error (MAE), and the number of parameters.

Network	PSNR (mean $\pm$ var)	MAE (mean $\pm$ var)	No. of Parameters
Unfolded Proximal Gradient Network	23.37 ± 2.6	$4.3 \times 10^{-2} \pm 4.27 \times 10^{-5}$	~ 175 k
UNet Lite	23.1 ± 4.21	$4.3 \times 10^{-2} \pm 8.2 \times 10^{-5}$	~ 160 k
UNet	20.21 ± 1.85	$5.14 \times 10^{-2} \pm 6.92 \times 10^{-5}$	~ 1175 k
ResNet	21.42 ± 1.21	$7.03 \times 10^{-2} \pm 6.94 \times 10^{-5}$	~ 226 k

with Leaky ReLU activations, which is then learnt from the training data.

Rather than crafting an analytical form for the regulariser, we aim to learn this directly from our training data. Thus, we may unfold the iterative algorithm (4), into a K-layered Neural Network denoted by U_θ [13, 14, 15], as shown in Fig 1. This allows us to learn all available free parameters from data. Consequently, we are also able to avoid the computational ambiguity of iterative algorithms, and fix the computational complexity of the solution.

3. DATA ACQUISITION

The dataset used in this study consists of six different recordings performed during IVUS-catheter pullbacks in a porcine model, that were acquired apriori. We then randomly sample 150 images from each of these pullbacks to obtain 900 frames in total, of which 150 frames are assigned to the validation set. The pullbacks are done at approximately 1 mm/s for a length 30 mm. The data is recorded in raw channel RF format. After beamforming, each frame contains 456 scanlines, with 520 samples along the penetration depth of 8 mm.

The training data is under-sampled by a factor of 4, by only considering every fourth receiving transducer element (from a 114 element transducer array). After beamforming and log-compression, we convert all frames to the polar domain and train the networks. We may then transform the predicted result back to the cartesian domain for display.

4. TRAINING STRATEGY

We employ an adversarial training strategy to train our network. In this case, the role of the adversary is fulfilled by a four layered convolutional neural network configured as a PatchGAN [16].

We use a combination of pixel based loss term (ℓ_1), denoted as $\mathcal{L}_{\ell_1}$, and a distribution-based loss term (adversarial loss; Binary Cross Entropy), denoted as $\mathcal{L}_{adv}$. In addition to these terms, we also use an ℓ_2 term between subsequent images ($(t-1)^{th}$ and t^{th} frame), so as to reduce uncorrelated noise in the predictions, and to promote consistency. This is given as,

$$\mathcal{L}_{ftf} = \|(U_\theta(\mathbf{y}_{t-1})) - (U_\theta(\mathbf{y}_t))\|_2^2 . \tag{8}$$

We then combine the individual loss terms into one loss function, defined as:

$$\mathcal{L}_{tot} = \lambda_1 \cdot \mathcal{L}_{adv} + \lambda_2 \cdot \mathcal{L}_{\ell_1} + \lambda_3 \cdot \mathcal{L}_{ftf}, \tag{9}$$

where λ_i are weight terms, with $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, and $\lambda_3 = 0.001$. All λ_i terms were determined empirically.

Furthermore, we train the network in a greedy manner [17]. In effect, we train the network one fold at a time, in conjunction with the K^{th} fold. We then continue to add intermediary folds, while freezing the previously trained fold. This process is continued until the full network is trained. Each step in the process is trained for 1500 epochs, which leads to a total of 6000 epochs, for a 5 fold network.

We use the Adam optimiser [18] with a learning rate of 10^{-5}, and $\beta_1 = 0.5$. All other optimiser parameters are left at default, as described in [18]. All of the networks were implemented using Python 3 and TensorFlow 2 [19], and trained using an NVIDIA GPU.

5. RESULTS

Fig. 2 displays a series of images including the input, the given ground truth, a prediction from the proposed network, and the prediction from a benchmark network, a UNet like (UNet Lite) neural network. Of the chosen benchmarks, the UNet Lite performed the best, which is why we have opted to highlight it. UNets are commonly used for comparable tasks, and due to its architecture that contains skip connections, it serves as a good common denominator among the most popular choices for our task. Furthermore, Fig. 2 shows highlighted areas within the image so that we may examine the quality of the prediction at a more detailed level.

In Fig. 2, we notice that both networks manage to reconstruct the ground truth to a large extent. However, we can visually verify the fact that the interface between tissues is much better defined with the proposed network, than the benchmark networks, in addition to reproducing a speckle

pattern that is qualitatively, much more representative of the ground truth.

Table 1 shows the comparison between the proposed method, and the benchmark networks. Although the Peak Signal to Noise Ratio (PSNR) values are quite close to each other, we do notice a bigger difference in the variance of the PSNR values between the two best performing networks. This makes a case for the consistency of the predictions produced by the proposed network. A similar trend can be observed with the Mean Absolute Error (MAE).

Neural networks are designed to optimise for the lowest value of the designated loss function rather than the best image, perceptually. Although the training process for all benchmarks were the same (albeit without the greedy training process), we see that it has not managed to optimise for the best perceptual image quality. This also alludes to why the test metrics may be so close to each other, even though the results can be discriminated from each other upon visual inspection.

6. CONCLUSION

IVUS is a valuable imaging modality supporting in the diagnosis and treatment of multiple coronary diseases. However, due to its limited bandwidth, it requires finding an optimal balance between spatial and temporal resolution. To address this issue, we have presented a model-based neural architecture that aims to reconstruct fully-sampled images, from under-sampled data. Based on the results presented in Fig. 2 and Table 1, we can conclude that the proposed network outperforms the benchmarks in reconstructing the given ground truth. The result is obtained with a reduction in the required data by a factor of 4.

While this is an encouraging result, it represents only the the first step along our research line. Exploring the augmentation of the training data with *in-silico* samples, and a learned model-based regulariser will be crucial next steps in improving the results from this baseline.

7. REFERENCES

[1] K Wayne Johnston, Robert B Rutherford, M David Tilson, Dhiraj M Shah, Larry Hollier, James C Stanley, et al., "Suggested standards for reporting on arterial aneurysms," *Journal of vascular surgery*, vol. 13, no. 3, pp. 452–458, 1991.

[2] Hector M Garcia-Garcia, Marco A Costa, and Patrick W Serruys, "Imaging of coronary atherosclerosis: intravascular ultrasound," *European heart journal*, vol. 31, no. 20, pp. 2456–2469, 2010.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[4] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[5] Yankun Cao, Ziqiao Wang, Zhi Liu, Yujun Li, Xiaoyan Xiao, Longkun Sun, Yang Zhang, Haixia Hou, Pengfei Zhang, and Guang Yang, "Multiparameter synchronous measurement with ivus images for intelligently diagnosing coronary cardiac disease," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[6] Shengran Su, Zhifan Gao, Heye Zhang, Qiang Lin, William Kongto Hau, and Shuo Li, "Detection of lumen and media-adventitia borders in ivus images using sparse auto-encoder neural network," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 1120–1124.

[7] Ji Yang, Lin Tong, Mehdi Faraji, and Anup Basu, "Ivus-net: an intravascular ultrasound segmentation network," in *International Conference on Smart Multimedia*. Springer, 2018, pp. 367–377.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[9] Joseph Paul Cohen, Margaux Luck, and Sina Honari, "Distribution matching losses can hallucinate features in medical image translation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 529–536.

[10] Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D Hoffman, et al., "Underspecification presents challenges for credibility in modern machine learning," *arXiv preprint arXiv:2011.03395*, 2020.

[11] Nishith Chennakeshava, Ben Luijten, Oded Drori, Massimo Mischi, Yonina C Eldar, and Ruud JG van Sloun, "High resolution plane wave compounding through deep proximal learning," in *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2020, pp. 1–4.

[12] Ruud JG van Sloun, Regev Cohen, and Yonina C Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE 108 (1)*, pp. 11–29, 2019.

[13] Karol Gregor and Yann LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 399–406.

[14] Vishal Monga, Yuelong Li, and Yonina C Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *arXiv preprint arXiv:1912.10557*, 2019.

[15] Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud JG van Sloun, and Yonina C Eldar, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE transactions on medical imaging 39(4)*, pp. 1051–1063, 2019.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[17] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al., "Greedy layer-wise training of deep networks," *Advances in neural information processing systems*, vol. 19, pp. 153, 2007.

[18] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, Software available from tensorflow.org.