Program Digest, 2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

MimickNet, Matching Clinical Post-Processing Under Realistic Black-Box Constraints

Ouwen Huang, Will Long, Nick Bottenus, Gregg E. Trahey, Sina Farsiu, Mark L. Palmeri
Department of Biomedical Engineering, Duke University, Durham, USA
Email: *ouwen.huang@duke.edu*

***Abstract*—Image post-processing is used in clinical-grade ultrasound scanners to improve image quality (e.g., reduce speckle noise and enhance contrast). These post-processing techniques vary across manufacturers and are generally kept proprietary, which presents a challenge for researchers looking to match current clinical-grade workflows. We introduce a deep learning framework, MimickNet, that transforms conventional delay-and-summed (DAS) beamformed images into the approximate post-processed images found on clinical-grade scanners. Training MimickNet only requires post-processed image samples from a scanner of interest without the need for explicit pairing to DAS data. Unpaired image flexibility allows MimckNet to hypothetically approximate any manufacturer's post-processing without hacking into commercial machines for pre-processed data. MimickNet generates images with an average similarity index measurement (SSIM) of 0.930±0.0892 on a 300 cineloop test set, and it generalizes to cardiac cineloops achieving an SSIM of 0.967±0.002 despite using no cardiac data in the training process. To our knowledge, this is the first work to approximate current clinical-grade ultrasound post-processing under realistic black-box constraints where before and after post-processing data is unavailable. MimickNet can be used out of the box or retrained to serve as a clinical post-processing baseline to compare against for future works in ultrasound image formation. To this end, we have made the MimickNet software open source at `https://github.com/ouwen/mimicknet`.**

***Index Terms*—MimickNet, Clinical Image Enhancement**

I. Introduction and Background

In the typical clinical B-mode ultrasound imaging paradigm, a transducer probe will transmit acoustic energy into tissue, and the back-scatter energy is reconstructed via beamforming techniques into a human eye-friendly image. This image attempts to faithfully map tissue's acoustic impedance, which is a property of its bulk modulus and density. Unfortunately, there are many sources of image degradation such as electronic noise, speckle from sub-resolution scatterers, reverberation, and de-focusing caused by heterogeneity in tissue sound speed [1]. In the literature, these sources of image degradation can be suppressed through better focusing [2], [3], spatial compounding [4], harmonic imaging [5], and coherence imaging techniques [6], [7].

In addition to beamforming, image post-processing is a significant contributor to image quality improvement. Reader studies have shown that medical providers largely prefer post-processed images over DAS beamformed imagery [8], [6]. Unfortunately, commercial post-processing algorithms are proprietary, and implementation details are typically kept as a black-box to the end-user. Thus, researchers that develop image improvement techniques on highly configurable research systems, such as Verasonics and Cephasonics scanners, face challenges in presenting their images alongside current clinical system scanner baselines. The current status quo for researchers working on novel image forming techniques is to compare against DAS beamformed data which is not typically viewed by medical providers. The optimal comparison would be pixel-wise to current clinical-grade standards, but researchers would either need access to proprietary post-processing code or access to unprocessed DAS data from difficult-to-configure commercial scanners. We aim to remove

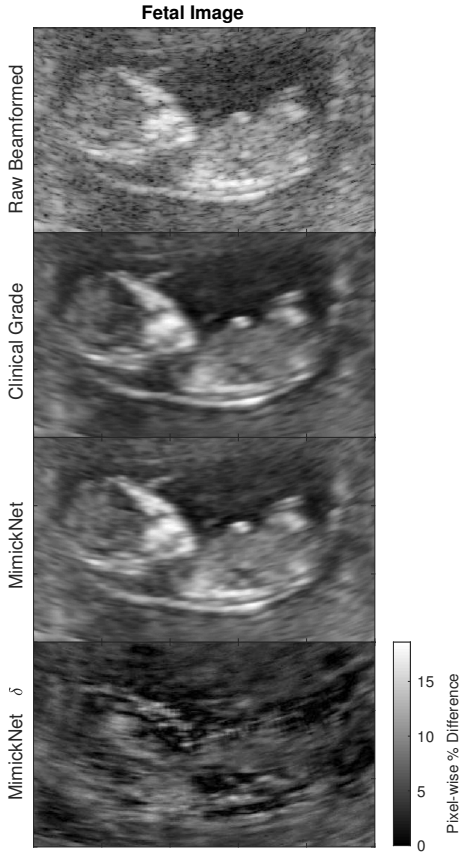

Fig. 1. Fetal image comparing clinical-grade post-processed images (ground truth) and MimickNet post-processing. In the last row, the difference between clinical-grade and MimickNet post-processing is scaled to maximize dynamic range. The SSIM of the MimickNet image to clinical grade image is 0.972

these barriers by leveraging recent generative adversarial network (GAN) methods developed in the deep learning field [9].

GANs use an adversarial objective function which are a unique class of distance functions that have shown success in the related field of image generation [10]. The adversarial objective optimizes two networks simultaneously. Given training batch sizes of m with individual examples i, G is a network that generates images from noise $z^{(i)}$, and another network, D, discriminates between real images $x^{(i)}$ and fake generated images $G(z^{(i)})$. D and G play a min-max game since they have competing objective functions shown in Eq. 1 and Eq. 2 where θ_g are parameters of G and θ_d are parameters of D. If this min-max game converges, G ultimately learns to generate realistic fake images that are indistinguishable from the perspective of D.

$$\underset{\theta_g}{\operatorname{argmin}} \frac{1}{m}\sum_{i=1}^{m}\log(1 - D_{\theta_d}(G_{\theta_g}(z^{(i)}))), \tag{1}$$

$$\underset{\theta_d}{\operatorname{argmax}} \frac{1}{m}\sum_{i=1}^{m}\log D_{\theta_d}(x^{(i)}) + \log(1 - D_{\theta_d}(G_{\theta_g}(z^{(i)}))). \tag{2}$$

Conditional GANs (cGANs) have seen success in image restoration as well as style transfer. With cGANs, a structured input, such as an image segmentation or corrupted image, is given instead of random noise [11].

In the field of ultrasound, deep learning techniques using cGANs and Unet convolutional neural networks (CNNs) [12] have recently been applied to B-mode imaging. Using CNNs it is possible to take a sequence of low resolution ultrasound images to construct a high resolution image [13]. CNNs and GANs have both been used to improve quality of plane wave reconstructions [14], [15]. GANs have been used to create a fast approximation of the known speckle reduction algorithm non-local low-rank (NLLR) [16]. Finally, GANs have been shown to have the ability to directly beam-form images [17], [18]. These and future novel image formation algorithms would benefit from comparing to a pixel-wise clinical-grade baseline. Unfortunately, this is a luxury not often available in most research environments.

We are interested in using an extension of GANs known as cycle-consistent GANs (CycleGAN) [19] to approximate clinical-grade post-processing which does not require training data to be registered together. This unique approach would allow us to approximate current day commercial algorithms with data simply acquired through a scanner's intended use.

CycleGANs consist of two key components: forward-reverse domain generators, G_a and G_b, and forward-reverse domain discriminators, D_a and D_b. The generators translate images from one domain to another, and the discriminators distinguish between real and fake generated images in each domain. We show the objective functions for one direction of the cycle in Eq. 3 and Eq. 4 where $a^{(i)}$ is an image from domain A, and $b^{(i)}$ is an image from domain B. In Eq. 3 and Eq. 4, the variables θ_{g_a} and θ_{d_b} are the parameters for the domain A forward generator and domain B discriminator. In Eq. 3, f can represent any distance metric to compare two images.

$$\underset{\theta_{g_a}}{\operatorname{argmin}}\ f(G_a(G_b(a^{(i)})), a^{(i)}), \tag{3}$$

$$\underset{\theta_{d_b}}{\operatorname{argmax}} \frac{1}{m}\sum_{i=1}^{m}\log D_b(b^{(i)}) + \log(1 - D_b(G_b(a^{(i)}))). \tag{4}$$

In this work, we investigate if it is possible to approximate post-processing algorithms found on clinical-grade scanners given DAS beamformed data as input to CNN generators. We first show what is theoretically feasible when before and after image pairs are provided and refer to this as a gray-box constraint. We view this as the classic image restoration problem where clinical-grade post-processed images are ground truth, and DAS data are "corrupted". Later, we constrain ourselves to the more realistic black-box setting where no before and after image pairs are available. We view this problem from the style transfer lens and train a CycleGAN from scratch to mimic clinical-grade post-processing. We refer to this trained model configuration as MimickNet. Our results suggest that any manufacturers' post-processing can be well approximated using this framework with just data acquired through a clinical scanner's intended use.

II. Methodology

TABLE I
DATASET OVERVIEW

Scanner Type	Targets	Frames	Train Frames	Test Frames
S2000	873	3085	2543	542
SC2000	158	12806	9754	3052
Verasonics	469	23309	18394	4915
Total	**1500**	**39200**	**30691**	**8509**

We start with 1500 unique ultrasound image cineloops from fetal, phantom, and liver targets across Siemens S2000, SC2000, or Verasonics Vantage scanners using various scan parameters from [20], [21], [6], [22]. This study was approved by the IRB at the Duke University, and each study subject provided written informed consent prior to enrollment in the study. We split whole cineloops into respective training and testing sets. Each cineloop has multiple image frames of conventional delay and summed (DAS) beamformed data. The datasets combined consist of 39200 frames with a 30691/8509 image frame train-test split. Each image frame runs through a Siemens proprietary compiled post-processing software producing before and after pairs. These pairs are shuffled and randomly cropped to 512x512 images with padded reflection if the dimensions are too small. Constraining the image dimensions enables batch training, which leads to faster and more stable training convergence. During inference time, images can be any size as long as they are divisible by 16 due to required padding in our CNN architecture. Table I contains details about our training data.

A. Gray-box Performance with Paired Images

In the gray-box case where before and after paired images are available, our problem can be seen as a classic image restoration problem where our input DAS beamformed data is "corrupted", and our clinical-grade post-processed image is the "uncorrupted ground truth". We optimize for the different distance metrics MSE, MAE, and SSIM. As defined in Eq. 5, MSE is the summed pixel-wise squared difference between a ground truth pixel $y^{(i)}$ in image y and estimated pixel $x^{(i)}$ in image x. These residuals are averaged by all pixels m in the image. MAE is defined in Eq. 6 as the summed pixel-wise absolute difference. SSIM is defined in Eq. 7 and is the multiplicative similarity between two images' luminance l, contrast c, and structure s (Eq. 8-10). X and Y define 11×11 kernels on two images. These kernels slide across the two images, and the output values are averaged to get the SSIM between two images. Variables μ_X, σ_X^2 and μ_Y, σ_Y^2 are the mean and variance of each kernel patch, respectively. Variables c_1, c_2, and c_3 are the constants $(k_1L)^2$, $(k_2L)^2$, and $c_2/2$ respectively. L is the dynamic range of the two images, k_1 is 0.01, and k_2 is 0.03. SSIM constants we use are based on [23].

$$MSE(x,y) = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - y^{(i)})^2, \tag{5}$$

$$MAE(x,y) = \frac{1}{m}\sum_{i=1}^{m}|x^{(i)} - y^{(i)}|, \tag{6}$$

$$SSIM(X,Y) = l(X,Y) * c(X,Y) * s(X,Y), \tag{7}$$

$$l(X,Y) = \frac{2\mu_X\mu_Y + c_1}{\mu_X^2 + \mu_Y^2 + c_1}, \tag{8}$$

$$c(X,Y) = \frac{2\sigma_X\sigma_Y + c_2}{\sigma_X^2 + \sigma_Y^2 + c_2}, \tag{9}$$

$$s(X,Y) = \frac{\sigma_{XY} + c_3}{\sigma_X\sigma_Y + c_3}. \tag{10}$$

B. Black-box Performance with Unpaired Images

To simulate the more realistic black-box case where paired before and after images are unavailable, we take whole cineloops from the training set used in the gray-box case and split them into two groups. For the first group, we only use the DAS beamformed data, and for the second group, we only use the clinical-grade post-processed data. We then train a CycleGAN using different distance metrics MSE, MAE, and SSIM for our generators' cycle-consistency loss (Eq. 3). Like in the gray-box case, MSE, MAE, and SSIM metrics were calculated by running our trained model on the full test set to their original non-padded size. Since we have access to the underlying proprietary clinical post-processing, we can compare against objective ground truths solely for final evaluation.

C. Generator and Discriminator Structure

The same overall generator network structure is used in both the gray-box and black-box cases. We use a simple encoder-decoder with skip connections as seen on the left side of Fig. 2. We vary filter sizes and the number of filters per layer as hyperparameters to the generator, and we report the total number of weight parameters in each model variation.

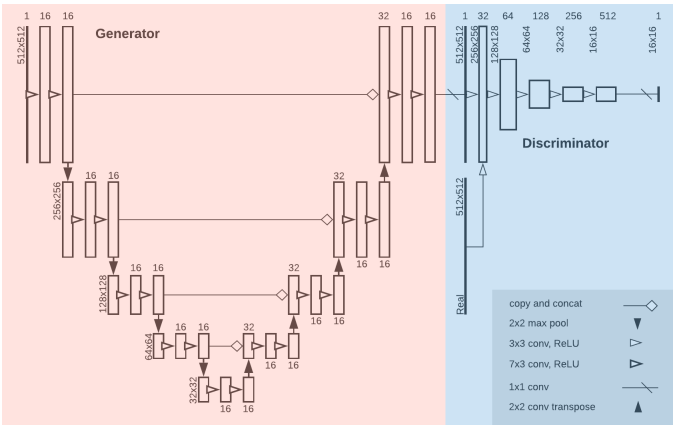

Fig. 2. Above is a diagram of the generator and discriminator structure for MimickNet in one translation direction. Note: the reverse translation direction uses an identical mirrored structure. Under gray-box training constraints, only the generator is used.

The discriminator structure on the right side of Fig. 2 follows the PatchGAN and LSGAN approach used in [11], [24] to optimize for least-squares on patches of linearly activated final outputs. The discriminator is only used to facilitate training in the black-box case where no paired images are available, and it is not used in the gray-box case since ground truths are available.

D. Worst Case Performance

$$cs(X,Y) = \frac{2\sigma_{XY} + c_2}{\sigma_X^2 + \sigma_Y^2 + c_2}. \tag{11}$$

We investigate outlier images that perform worst on the SSIM metric by breaking SSIM into its three components: luminance l, contrast c, and structure s. The equations for contrast c and structure s are highly related in examining variance between and within patches. Thus, c (Eq. 9) and s (Eq. 10) are multiplied together into a single contrast-structure cs equation (Eq. 11).

III. Results

A. Gray-Box Performance with Paired Images

In the theoretical gray-box case where before and after paired images are available, we explore different possible Unet encoder-decoder hyperparameters. For each hyperparameter variation, we trained a triplet of models that optimize for SSIM, MSE, and MAE. We note that within each triplet, models using the SSIM minimization objective have the best SSIM. We are primarily interested in the best SSIM metric since it was originally formulated to model the human visual system [23] and provides the best dynamic range for quality on our dataset. In Table II, many of the metrics across model variations are not significantly different, but regardless achieve a high SSIM.

TABLE II
GRAY-BOX PERFORMANCE WITH PAIRED IMAGES

Loss	**Params**	**MSE** 10^{-3}	**MAE** 10^{-2}	**SSIM**
ssim	34849	2.49±2.88	3.78±2.28	0.975±0.013
mse	34849	2.27±2.41	3.67±1.92	0.950±0.019
mae	34849	2.31±2.54	3.68±1.96	0.951±0.016
ssim	52993	2.28±2.77	3.65±2.24	0.979±0.013
mse	52993	2.19±2.40	3.60±1.92	0.956±0.017
mae	52993	2.11±2.35	3.52±1.89	0.959±0.015
ssim	77185	2.38±2.91	3.70±2.28	0.976±0.015
mse	77185	2.02±2.09	3.46±1.70	0.946±0.022
mae	77185	2.14±2.23	3.55±1.80	0.947±0.020
ssim	117697	2.22±2.65	3.59±2.11	0.977±0.014
mse	117697	2.72±2.51	4.07±1.95	0.931±0.023
mae	117697	2.93±2.93	4.18±2.11	0.927±0.022

B. Black-box Performance with Unpaired Images

In the more realistic black-box case where before and after images are not available, we also explore different Unet architecture hyperparameters. We attempted to train from scratch the same 52993 parameter generator network architecture selected from Table II, but we were unsuccessful in guiding convergence without increasing the number of generator parameters to 117697. This increase was accomplished by changing every filter size from 3×3 to 7×3, and metrics can be seen in Table III. For the large 7.76M parameter generator network, performance differences between triplets of the objective functions are not significant.

We select the 117697 parameter network optimizing MSE for subsequent analysis since it achieves the highest SSIM with fewest parameters. We refer to this configuration, shown in Fig. 2, as MimickNet. In Fig. 1 and Fig. 3, fetal, liver, and phantom images are shown. Without the scaled absolute differences in the last rows, it is much more difficult to discern localized differences between MimickNet images and clinical-grade post-processed images.

TABLE III
BLACK-BOX PERFORMANCE WITH UNPAIRED IMAGES

Loss	**Params**	**MSE** 10^{-3}	**MAE** 10^{-2}	**SSIM**
ssim	117697	7.26±10.5	6.54±4.38	0.883±0.091
mse	117697	6.83±11.1	6.31±4.39	**0.930±0.089**
mae	117697	6.79±9.89	6.30±4.27	0.900±0.085
ssim	7.76M	**4.45±5.71**	**5.14±3.12**	0.918±0.078
mse	7.76M	6.23±6.30	6.14±3.24	0.897±0.052
mae	7.76M	6.20±9.10	6.02±4.21	0.918±0.084

C. Runtime Performance

In Table IV, the runtime was examined for the best SSIM performing model in the gray-box paired image and black-box unpaired image training cases. Frames per second (FPS) measurements were calculated for an NVIDIA P100. Floating-point operations per second (FLOPS) are provided as a hardware independent measurement since runtime generally scales linearly with the number of FLOPS used by the model. As a reference point, we include metrics from MobileNetV2 [25], a lightweight image classifier designed explicitly for use on mobile phones. MimickNet uses 2000x fewer FLOPS compared to MobileNetV2. Note that FPS measurements for MobileNetV2 were performed on a Google Pixel 1 phone from [25] and not an NVIDIA P100.

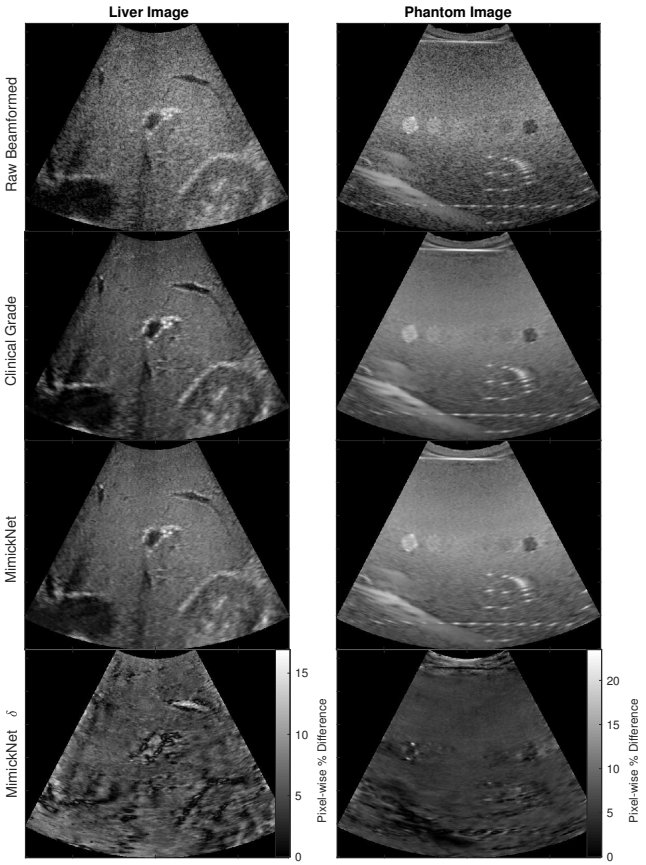

Fig. 3. Liver (left) and phantom (right) images. The difference between clinical-grade and MimickNet outputs are scaled to maximize dynamic range. The SSIM between MimickNet and clinical-grade images for the liver target is 0.9472. The SSIM between MimickNet and clinical-grade images for the phantom target is 0.9802

TABLE IV
RUNTIME PERFORMANCE ON NVIDIA P100 AND *PIXEL 1 PHONE UNDER GRAY-BOX AND BLACK-BOX TRAINING CONSTRAINTS

Model	Input Size	Params	MFLOPS	FPS (Hz)
Gray-box	512x512	52993	0.105	142
Black-box (MimickNet)	512x512	117697	0.235	92
MobileNetV2	224x224	4.3M	569	5*

D. Worst Case Performance

We investigate the distribution of SSIM across our entire test dataset. We break the the SSIM into its luminance l and contrast-structure cs components following Eq. 8 and Eq. 11. In Fig. 4, these components' histogram and kernel density estimate are plotted for the gray-box paired image and the black-box unpaired image training cases. The min-max cs range for the gray-box case is tightly between 0.950 and 0.998,

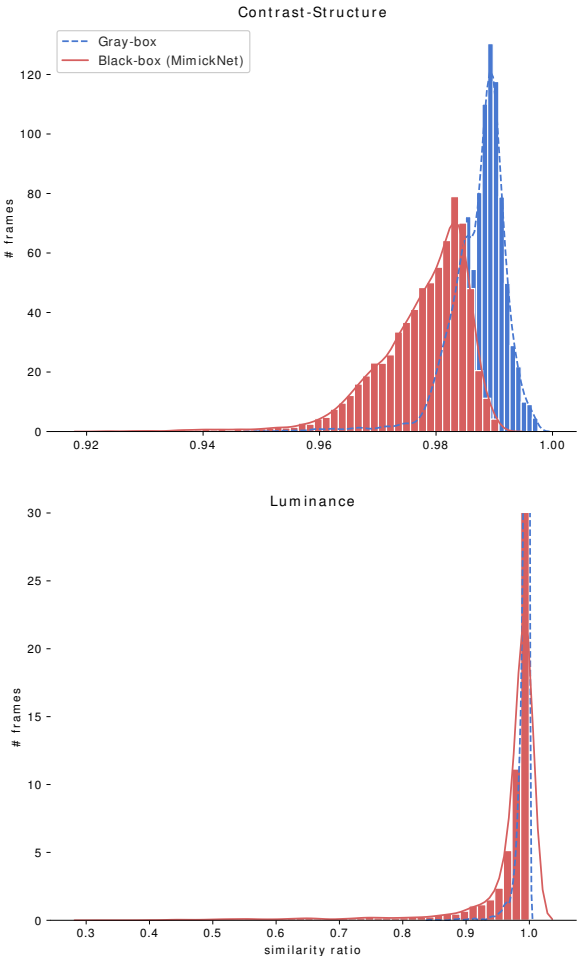

Fig. 4. The distribution of contrast-structure (top) and luminance (bottom) of all image frames in our test dataset produced under gray-box and black-box constraints. The cs is 0.987 ± 0.005 and l is 0.993 ± 0.0103 under gray-box constraints. The cs is 0.978 ± 0.008 and l is 0.967 ± 0.073 under black-box constraints.

and the black-box case overlaps this region with a min-max cs range between 0.922 and 0.990. The min-max l range of the gray-box case falls between 0.842 and 1.000, but the black-box case has a large min-max range of 0.318 and 1.000.

We also closely investigated outlier images that perform poorly on the SSIM metric by looking at the worst images. Fig. 6 contains three representative images. We included gray-box image results to showcase better the performance gap between what is possible when paired images are available versus when they are not. All three images produced with black-box constraints have high contrast-structure cs, but variable luminance l.

E. Out of Dataset Distribution Performance

To assess the generalizability of MimickNet post-processing, we applied it to cardiac cineloop data. These data are outside of our train-test dataset distribution which only included phantom, fetal, and liver imaging targets. We also applied MimickNet post-processing to a recent novel beamforming method known as REFocUS [3] instead of

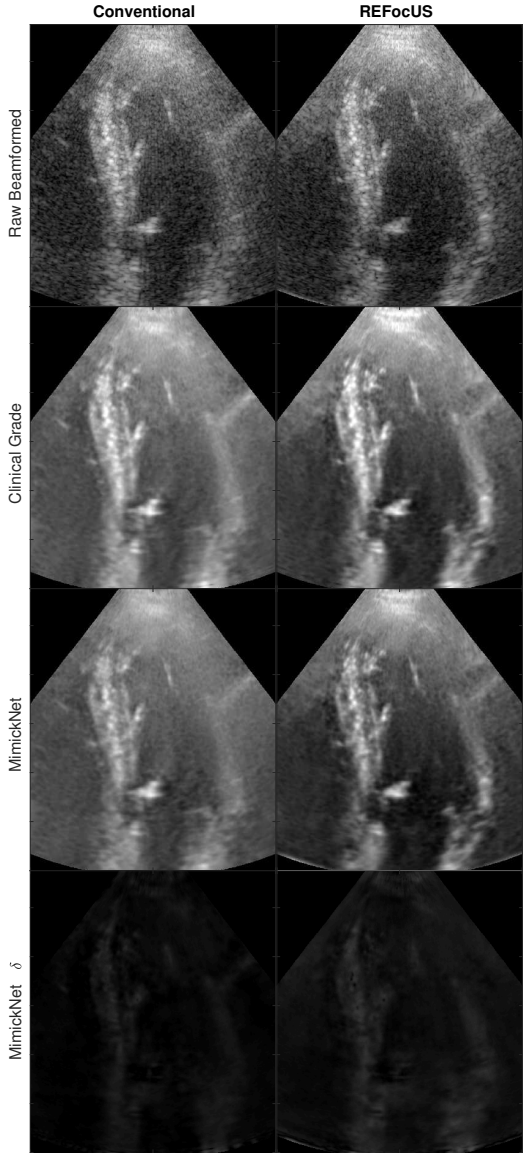

Fig. 5. MimickNet applied to out of distribution cardiac data on conventional dynamic receive images and REFocUS ultrasound beamformed images. MimickNet is only trained on fetal, liver, and phantom data. SSIM between clinical-grade post-processing and MimickNet for conventional DAS beamformed images and REFocUS beamformed images was 0.967 ± 0.002 and 0.950 ± 0.0157, respectively. The last row is at the same scale as the cardiac images above.

DAS images. REFocUS allows for transmit-receive focusing everywhere under linear system assumptions resulting in better image resolution and contrast-to-noise ratio. In Fig. 5, we see that MimickNet post-processed images closely match clinical-grade post-processing for conventional dynamic receive beamforming with an SSIM of 0.967 ± 0.002. Similar to clinical-grade post-processing, we see that contrast improvements in the heart chamber and resolution improvements along the heart septum due to REFocUS are preserved after MimickNet post-processing, achieving an SSIM of 0.950 ± 0.0157.

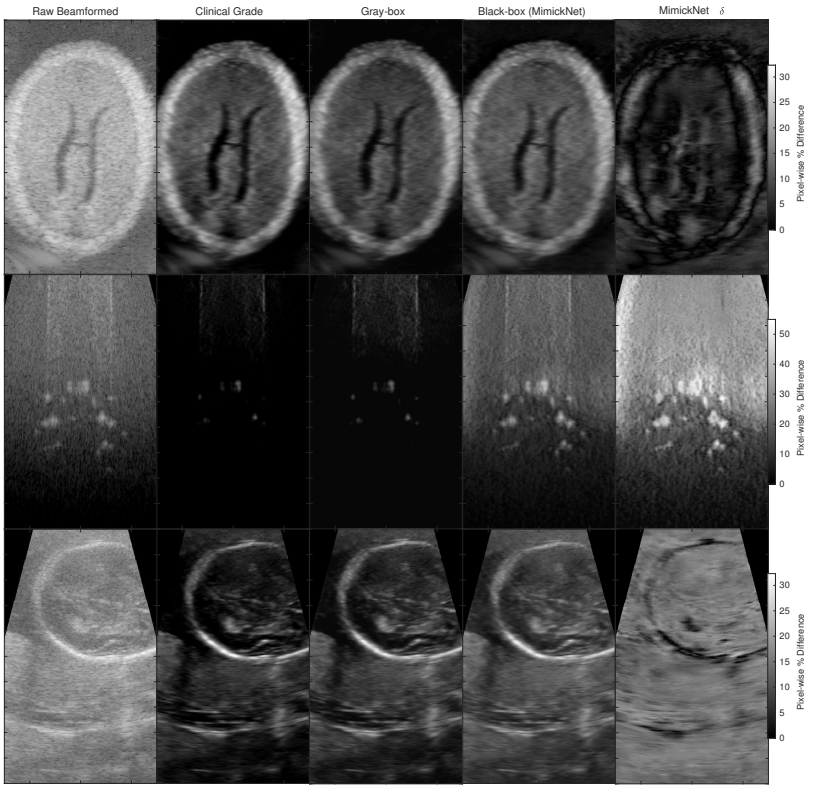

Fig. 6. The worst case scenario images for two fetal brain images (top, bottom), and a phantom (middle). The SSIM of the black-box case (MimickNet) to the ground truth images from top to bottom is 0.665 (cs = 0.962, l=0.681), 0.414 (cs = 0.947, l = 0.419), and 0.603 (cs = 0.964, l = 0.612). The SSIM of the gray-box case to the ground truth images from top to bottom is 0.873 (cs=0.984, l=0.883), 0.967 (cs=0.996, l=0.971), and 0.901 (cs=0.988, l=0.911). Here l is the luminance and cs is the contrast-structure components of SSIM.

IV. Discussion

MimickNet can closely approximate clinical-grade post-processing with an SSIM of 0.930±0.089 such that even upon close inspection, few differences are observed. This performance was achieved without knowledge of the pre-processed pair. We do observe a performance gap compared to the gray-box setting, which achieves an SSIM of 0.979±0.013. However, emulating the gray-box setting would require researchers to tamper with scanner systems to siphon off pre-processed data, so we explore ways to eliminate this gap.

The performance gap is primarily attributed to differences in image luminance from outlier frames seen in Fig. 4. Although images generated under black-box constraints present a large min-max l range of 0.318 to 1.000, we note that the mean and standard deviation is 0.967 ± 0.073. Therefore, the majority of images do have well-approximated luminance, despite the sizeable min-max range. For the two fetal brain images in Fig. 6, we qualitatively see that much of the contrast and structure are preserved while luminance is not. This matches the measured quantitative cs and l SSIM components.

We found it interesting that clinical-grade post-processing would remove such bright reflectors seen in the DAS beamformed phantom image (Fig. 6, 2nd row). This level of artifact removal likely requires window clipping. When we clip the lower dynamic range of DAS beamformed data from -120dB to -80dB, we see the bright scatterers in DAS beamformed images dim and practically match clinical-grade post-processing without any additional changes. Conceptually, clipping values to -80dB is a reasonable choice since it is close to the noise floor of most ultrasound transducers. In the CycleGAN training paradigm, it can be challenging to learn these clipping cutoffs due to the cycle-consistency loss (defined in Eq. 3). The backward generator would be penalized by any information destroyed through clipping learned in the forward generator. Since the cycle-consistency loss does not exist in optimization under the gray-box setting, the model under gray-box settings can learn the clipping better than under black-box settings. Fortunately, luminance can be modified to a large extent in real-time by changing the imaging window or gain by ultrasound end-users.

As is, MimickNet shows promise for production use. It runs in real-time at 92 FPS on an NVIDIA P100 and uses 2000x fewer FLOPS than models such as MobileNetV2, which was designed for less capable hardware such as mobile phone CPUs. This runtime is relevant since more ultrasound systems are being developed for mobile phone viewing [26]. Future work will assess the performance of MimickNet on mobile phones and other data or compute constrained settings.

This work's main contribution is in decreasing the barrier of clinical translation for future research. Medical images previously only understood by research domain experts can be translated to clinical-grade images widely familiar to medical providers. Future work will aim to implement a flexible end-to-end software package to train a mimic provided data from two arbitrary scanner systems. Future work will also examine how much data is required to create a high-performance mimic. Our results with unpaired domain translation suggest a similar method could be used to approximate medical image post-processing in other modalities such as CT and MR.

V. Conclusion

MimickNet closely approximates current clinical post-processing in the realistic black-box setting where before and after post-processing image pairs are unavailable. We present it as an image matching tool to provide fair comparisons of novel beamforming and image formation techniques to a clinical baseline mimic. It runs in real-time, works for out-of-distribution cardiac data, and thus shows promise for practical production use. We demonstrated its application in comparing different beamforming methods with clinical-grade post-processing and showed that resolution improvements are carried over into the final post-processed image.

Acknowledgment

This work was supported by the National Institute of Biomedical Imaging and Bioengineering under Grant R01-EB026574, and National Institutes of Health under Grant 5T32GM007171-44. The authors would like to thank Siemens Medical Inc. USA for in kind technical support.

References

[1] G. F. Pinton, G. E. Trahey, and J. J. Dahl, "Sources of image degradation in fundamental and harmonic ultrasound imaging using nonlinear, full-wave simulations," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 58, no. 4, pp. 754–765, Apr. 2011.

[2] K. Thiele, J. Jago, R. Entrekin, and R. Peterson, "Exploring nsight imaging, a totally new architecture for premium ultrasound," Philips, Tech. Rep. 4522 962 95791, June 2013. [Online]. Available: https://www.usa.philips.com/healthcare/resources/feature-detail/nsight

[3] N. Bottenus, "REFoCUS: Ultrasound focusing for the software beamforming age," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2018, pp. 1–4.

[4] G. E. Trahey, J. W. Allison, S. W. Smith, and O. T. von Ramm, "Speckle reduction achievable by spatial compounding and frequency compounding: Experimental results and implications for target detectability," in *Pattern Recognition and Acoustical Imaging*, vol. 0768. International Society for Optics and Photonics, Sep. 1987, pp. 185–192.

[5] A. Anvari, F. Forsberg, and A. E. Samir, "A primer on the physical principles of tissue harmonic imaging," *Radiographics*, vol. 35, no. 7, pp. 1955–1964, Nov. 2015.

[6] W. Long, D. Hyun, K. R. Choudhury, D. Bradway, P. McNally, B. Boyd, S. Ellestad, and G. E. Trahey, "Clinical utility of fetal Short-Lag spatial coherence imaging," *Ultrasound Med. Biol.*, vol. 44, no. 4, pp. 794–806, Apr. 2018.

[7] M. R. Morgan, D. Hyun, and G. E. Trahey, "Short-lag spatial coherence imaging in 1.5-d and 1.75-d arrays: Elevation performance and array design considerations," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, Mar. 2019.

[8] H. Ahman, L. Thompson, A. Swarbrick, and J. Woodward, "Understanding the advanced signal processing technique of Real-Time adaptive filters," *J. Diagn. Med. Sonogr.*, vol. 25, no. 3, pp. 145–160, May 2009.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680.

[10] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=B1xsqj09Fm

[11] P. Isola, J.-Y. Zhu, T. Zhou, and A. 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.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Med. Image Comput. Comput. Assist. Interv.*, 2015.

[13] M. Abdel-Nasser and O. A. Omer, "Ultrasound image enhancement using a deep learning architecture," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*. Springer International Publishing, 2017, pp. 639–649.

[14] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, and J.-P. Thiran, "Deep convolutional neural network for ultrasound image enhancement," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2018, pp. 1–4.

[15] X. Zhang, J. Li, Q. He, H. Zhang, and J. Luo, "High-quality reconstruction of plane-wave imaging using generative adversarial network," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct 2018, pp. 1–4.

[16] F. Dietrichson, E. Smistad, A. Ostvik, and L. Lovstakken, "Ultrasound speckle reduction using generative adversial networks," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2018, pp. 1–4.

[17] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell, "A generative adversarial neural network for beamforming ultrasound images : Invited presentation," in *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, March 2019, pp. 1–6.

[18] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910, May 2019.

[19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[20] V. Kakkad, J. Dahl, S. Ellestad, and G. Trahey, "In vivo application of short-lag spatial coherence and harmonic spatial coherence imaging in fetal ultrasound," *Ultrason. Imaging*, vol. 37, no. 2, pp. 101–116, Apr. 2015.

[21] Y. Deng, M. L. Palmeri, N. C. Rouze, G. E. Trahey, C. M. Haystead, and K. R. Nightingale, "Quantifying image quality improvement using elevated acoustic output in B-Mode harmonic imaging," *Ultrasound Med. Biol.*, vol. 43, no. 10, pp. 2416–2425, Oct. 2017.

[22] J. Long, W. Long, N. Bottenus, G. F. Pintonl, and G. E. Trahey, "Implications of lag-one coherence on real-time adaptive frequency selection," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–9.

[23] A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[24] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2813–2821.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 4510–4520.

[26] H. Hewener and S. Tretbar, "Mobile ultrafast ultrasound imaging system based on smartphone and tablet devices," in *2015 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2015, pp. 1–4.