Super-resolution Ultrasound imaging via Unpaired Training with the Model-Informed CycleGAN Algorithm

Vassili Pustovalov, Duong-Hung Pham, Denis Kouamé

SUPER-RESOLUTION ULTRASOUND IMAGING VIA UNPAIRED TRAINING WITH THE MODEL-INFORMED CYCLEGAN ALGORITHM

Vassili Pustovalov, Duong Hung Pham, Denis Kouamé

Université Toulouse III Paul Sabatier, IRIT, CNRS UMR 5505, Toulouse, France

ABSTRACT

In traditional ultrasound (US) imaging, there has long been a trade-off between spatial resolution and imaging frame rate. Super-resolution (SR) imaging techniques stand out as extensively studied methods to overcome the spatial resolution limitation. Recently, these techniques have garnered significant attention in ultrafast US imaging research, enabling enhanced visualization of microvasculature. However, current SR methods encounter challenges, including the need to acquire a substantial number of image frames over an extended acquisition time, coupled with intricate post-processing steps. Convolutional neural networks (CNNs) emerge as promising approaches, surpassing classical model-based methods. Nevertheless, they face challenges arising from the lack of *in vivo* data and the absence of ground truth. In this work, we introduce a *model-informed* CNN for computing SR images, addressing generalization issues through an unpaired and unlabeled training approach. Our results are compared with those obtained from state-of-the-art techniques on both simulated and *in vivo* data.

Index Terms— Super Resolution, Ultrasound Imaging, Microvasculature, CNN, Unpaired, Cyclic GAN

1. INTRODUCTION

Ultrafast US imaging has emerged as a prominent noninvasive tool for visualizing and quantifying microvascular structures with high precision [1]. It leverages advancements in US imaging including SR, leading to the emergence of new techniques such as functional US, or the enhancement of existing methods, exemplified by the precise detection of tumoral areas based on measurable microvascular properties [2]. Notably, the integration of this modality with intravascular microbubble (MB) contrast agents, serving as acoustic point sources at sub-diffraction precision, has recently introduced a novel SR imaging technique known as super-localization or ultrasound localization microscopy (ULM). Remarkably, this technique can achieve a revolutionary tenfold improvement in resolution [2]. Despite these continuous improvements, ULM still requires a prolonged data acquisition time and a sophisticated post-processing procedure.

CNNs-based methods offer an efficient alternative, providing enhanced performances with faster computation times. The advantage of these methods lies in their capacity to learn parameters from the data, enabling precise and rapid SR imaging. However, in the context of US imaging, CNN-based methods frequently encounter a significant limitation related to the insufficient availability of training data, especially the absence of ground truth (i.e. labeled by an expert or confirmed with optical imaging) for *in vivo* data. This challenge is sometimes addressed by utilizing the outcomes of optimization algorithms as ground truth after fine-tuning [3] or those achieved with the highest-performing US system [4]. Furthermore, a portion of the training is usually carried out on simulation data, with a trend towards using highly realistic simulations based on optical images of microvasculature and *in vivo* data [5]. Despite their potential, these approaches struggle to bridge the domain gap between simulation and *in vivo* data, thereby constraining the clinical applicability of CNN-based algorithms.

In this work, we introduce a specific CNN-based approach for achieving SR in US imaging, enabling rapid imaging of large vascular regions. This addresses the constraints related to the acquisition time, computation time, and data volume inherent in traditional ULM. Unlike existing CNN-based methods, our approach involves unpaired and unlabeled training, facilitated by the deployment of a Cyclic Generative Adversarial Network (CycleGAN) architecture. Notably, this architecture incorporates information on the explicit model of US image formation, resulting in a *model-informed* CNN design that enhances training efficiency and stability. The remainder of the paper is organized as follows. We begin by introducing the design of our CycleGAN architecture and its training on unpaired datasets in Section 2. Then, we demonstrate its capability to generate super-resolved images using *in vivo* data in Section 3. Finally, we conduct a comparative analysis between the performance of the proposed technique and state-of-the-art algorithms: mSPCN [6] and ULM [7].

2. METHODS

2.1. Model-informed CycleGAN Architecture

To tackle inverse problems, CNN-based methods typically learn the inverse mapping between input data and corresponding labeled data. However, in medical imaging, acquiring a sufficient quantity of matched label data for network train-

ing poses a significant challenge. Our strategy is to explore a CycleGAN, capable of learning from unmatched datasets. In this study, our objective is to perform SR from far fewer frames than ULM, thereby offering an alternative to super-localization-based techniques. Drawing inspiration from [5], which proposes an end-to-end CNN for super-resolved velocity map estimation without super-localization, our CycleGAN is trained to perform the SR task on temporal mean of tissue-filtered B-mode frames.

Our CycleGAN architecture, depicted in Fig. 1, comprises four independent neural networks that are trained simultaneously : a pair of two generators (G, SH) and two discriminators (D_X, D_Y). The generator G is a U-Net followed by two upsampling blocks, implementing an efficient sub-pixel convolution to achieve a fourfold SR, thereby mitigating gridding artifacts. This generator transforms low-resolution US images into high-resolution images. Meanwhile, the generator SH follows the forward model formulation (1) to produces low-resolution images from high-resolution ones. Both discriminators utilize a PatchGAN structure.

Training the CycleGAN presents a significant challenge due to the complexity of its architecture. To overcome this limitation, we embed information from the forward model of the inverse problem into the CycleGAN, leading to a *model-informed* CNN architecture. Specifically, we consider the following forward equation of imaging formulation:

$$Y = SHx + N, \tag{1}$$

where S represents the downsampling operator, H denotes the PSF, and N stands for an additive noise. In this context, the second generator replicates (1) with H being learned during the training. Consequently, the knowledge of the forward operator allows us to design a simple generator architecture that accelerates and stabilizes the training.

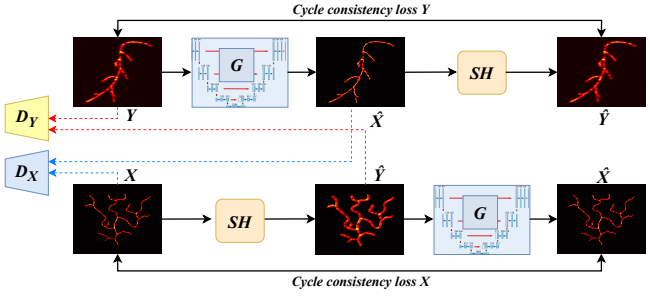


Fig. 1. CycleGAN architecture with two generators (*G*, *SH*) and two discriminators (D_Y, D_X). The *SH* network implements the direct model described by equation (1). *X* represents the high-resolution domain and *Y* represents the low-resolution domain. Datasets of domains *X* and *Y* are unpaired. Note that both G and SH are depicted twice for illustration purposes.

The stability of GAN training often depends on addressing the mode collapse problem. Mode collapse occurs when the generator network of the GAN produces highly similar outputs, indicating a failure to capture the entire distribution of the data, which is intricately spread out. In such cases, the generator generates only a small subset of the distribution. This issue arises when the discriminator network becomes excessively powerful compared to the generator, easily discerning the generated samples. Consequently, the discriminator provides feedback to generate samples very similar to the previous ones. To prevent our second generator from becoming trapped in a limited set of samples, we incorporate a random noise term into the input data. The intensity of this noise is learned during training.

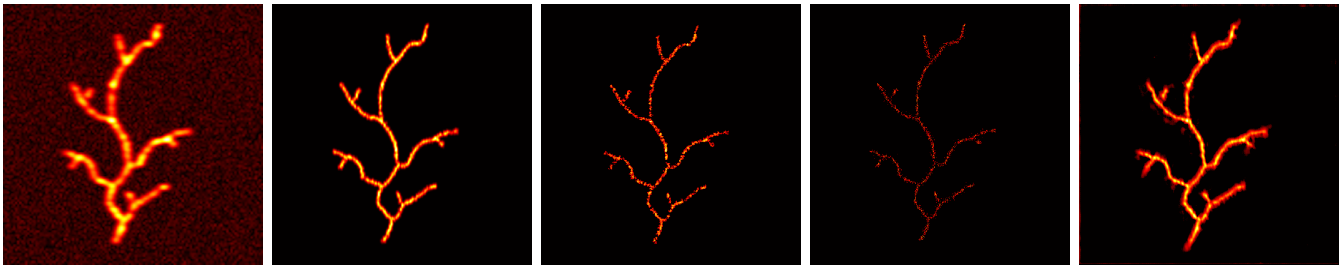
2.2. CycleGAN Loss Function

Recent CNNs utilizing CycleGAN [8] incorporate the Wasserstein distance, rooted in optimal transport theory, as a crucial element in the loss function. This distance metric serves as a robust measure between two probability distributions, namely the training data distribution (ground truth distribution) and the distribution of generated SR images. Nevertheless, Wasserstein GANs necessitate regularization techniques like weight clipping or gradient penalty to enforce the Lipschitz constraint on the discriminator networks [9]. While the preferred method is gradient penalty, it is often computationally expensive and tends to slow down the training process. In this study, we opted for discriminator training using the least squares loss (LS-GANs) as an equivalent alternative to the Wasserstein loss [8]. The least squares loss enhances training stability by reducing discriminator sensitivity to the generator's output. The cycle-consistency loss employs mean absolute errors to quantify the distance between the initial images and their reconstructions.

2.3. Training and Simulation Data

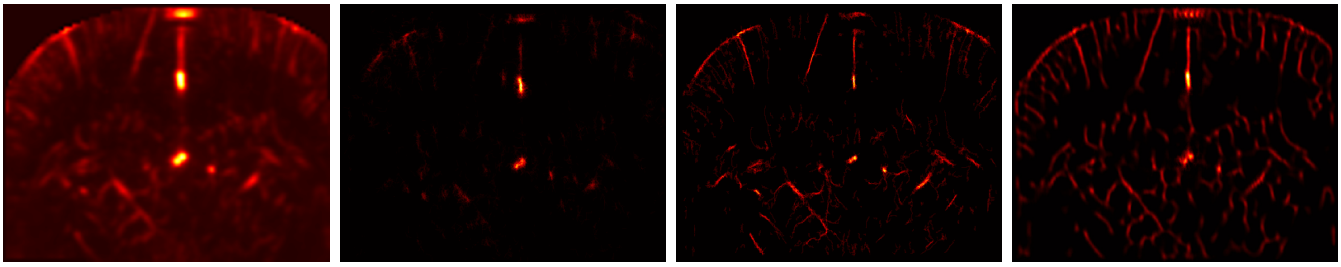
The CycleGAN was trained using a series of numerical simulations adapted from [10] (see Fig. 2), incorporating nine types of vascular structures mimicking *in vivo* blood vessels. To enhance the dataset's variability and address overfitting, new structures were meticulously designed. Specifically, we respectively employed 180 and 120 distinct simulated vascular structures to generate images from the high-resolution domain and from the low-resolution domain. An additional 8 structures were set aside for the validation set. To augment the dataset, we applied data augmentation techniques, including translations and rotations, to the generated images. In total, 6480 low-resolution images and 4320 high-resolution images were generated and used for training.

The CycleGAN was implemented using PyTorch and underwent training for 100 epochs employing the Adam optimizer, with a learning rate of 3×10^{-5} for the generators and 5×10^{-5} for the discriminators. The computational experiments were conducted on a high-performance server equipped with 32 GB RAM and 4 GTX-1080-TI GPUs.



(a) Temporal Mean (b) Ground Truth (c) mSPCN[6] (d) ULM [7] (e) CycleGAN (ours)

Fig. 2. Comparison of the super-resolved images obtained by mSPCN, and CycleGAN methods on numerical simulations: (a) Sum of Bmode filtered frames. (b) Ground Truth. (c) mSPCN-ULM[6] output. (d) ULM [7] output. (e) CycleGAN output.



(a) SVD-filtered (b) mSPCN[6] (c) ULM [7] (d) CycleGAN (ours)

Fig. 3. Comparison of the super resolved images obtained by mSPCN with tracking, classic ULM with tracking [7] and CycleGAN : (a) Sum of Bmode filtered frames. (b) mSPCN[6] output. (c) ULM output. (d) CycleGAN output.

3. RESULTS

3.1. Numerical Simulation Results

The simulation data used for comparison corresponds to the validation set of training data described in Subsection 2.3. We conducted a comparative analysis involving our CycleGAN and other techniques. The SR estimation results are depicted in Fig. 2. Fig. 2 (a) and (b) display the temporal mean of the B-mode frames and the corresponding ground truth, respectively. Fig. 2 (c) presents the SR estimation obtained with mSPCN [6]. Fig. 2 (d) corresponds to the result of ULM, while Fig. 2 (e) shows the SR estimation obtained with our CycleGAN. We perform a quantitative comparison of the results presented in Table 1 employing metrics consistent with those defined in [10], namely the MS-SSIM and the RSE. Additionally, we compare the GPU inference time, using a distinct set of simulation data, among various CNN methods. This includes a comparison with the method introduced in [10], which we will refer to as ULM-GAN[1]

By design, from each input B-mode frame, the mSPCN network and ULM produce a super-resolved image in which MBs are detected and super-localized. For performance comparison, the results of mSPCN and ULM are considered as the superimposition of all super-localized MBs across the B-mode sequence (i.e. the temporal sum of all the localized MBs). In contrast, the CycleGAN and ULM-GAN networks take the temporal mean of the US B-mode sequence displayed in Fig. 2 (a) as input which results in a single super-resolved image.

Table 1. MS-SSIM, RSE and CPU inference time comparison for the estimations of Fig. 2. The best results are highlighted in boldface.

	ULM-GAN [10]	mSPCN [6]	ULM [7]	CycleGAN
MS-SSIM	-	0.9211	0.9388	**0.9805**
RSE	-	0.1844	0.4453	**0.0946**
CPU Time (s)†	-	4.212	0.431	**0.405**
GPU Time (s)‡	2.09	2.63	-	**0.118**

† Average time over 100 samples computed with an Intel Core i5-10500 CPU @ 3.10GHz with 16 GB RAM on the simulated validation set.

‡ Average time over 100 samples computed using a Nvidia Tesla V100 with 16 GB RAM on simulation data adapted to resemble the test data from [10].

3.2. *In vivo* Results

In the *in vivo* data comparison, we utilize the initial block comprising 800 frames from the rat brain in the PALA dataset [11]. The pre-processing phase involves the application of the temporal and SVD filters to the US sequence to retain only

[1] Although our method outperforms ULM-GAN across all metrics on similar simulation data, we did not present the results, as it would not allow for a fair comparison due to the unavailability of the code and data from [10]. However, we opted to use their results for the comparison of GPU computation time only, as our network runs on the same GPU.

the MBs, thereby eliminating tissue and noise. Consistent filtering parameters are employed across all methods, ensuring parity in performance comparisons.

Similar to the approach applied to simulation data, both ULM and mSPCN undertake the processing of the SVD-filtered sequence, whereas CycleGAN takes the temporal mean image as input, as illustrated in Fig. 3 (a). Additionally, we exclude ULM-GAN from this *in vivo* comparison due to the unavailability of the code. It is noteworthy that in this *in vivo* experiment, we introduced the tracking step for ULM and mSPCN. This was introduced based on our observation that it contributed to the improvement of the final super-resolved image.

Since all the methods do not use the same metrics, we did not use the quantitative *in vivo* metrics used in ULM, such as the gridding index or the saturation. As a result, our comparison is restricted to the evaluation of execution times, as reported in Table 2, across various methods. Additionally, we conduct a qualitative analysis of the results, as in Fig. 3.

Table 2. Computation time comparison on CPU / GPU for the *in vivo* data. The best result is in boldface.

	mSPCN[6]	ULM [7]	CycleGAN
Time (s)	37.9 / 1.48	1.59 / -	**0.40 / 0.04**

3.3. Discussion

The experimental results reveal that using a short sequence as the input of super-localization techniques produces a SR image containing a limited number of MBs. Notably, some secondary blood vessels, visible on the temporal mean image in Fig. 3 (a), are only partially reconstructed by super-localization methods, resulting in an incomplete depiction of the microvasculature. In contrast, CycleGAN demonstrates the capability to directly generate a continuous super-resolved image from a short sequence, both in simulation and *in vivo* data. In quantitative assessments on simulation data, our method outperforms state of the art methods, yielding superior metrics. Our neural network demonstrates superior speed in both simulation and *in vivo* scenarios.

4. CONCLUSION

In this work we demonstrate the feasibility of obtaining SR-images from a limited number of low-resolution US frames, using the proposed *model-informed* CycleGAN training configuration. In contrast to existing super-localization techniques, our approach requires a lower acquisition time and provides robustness against artifacts such as motion artifacts. Moving forward, we aim to further refine this unpaired training method, focusing on re-implementing one of the ULM steps as a CNN.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using animal subjects data made available in open access in [11]. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

This research was supported by the National Research Agency (ANR) **SONATINE** project with grant number ANR-23-CE45-0002-01.

7. REFERENCES

[1] Mickael Tanter and Mathias Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE TUFFC*, vol. 61, no. 1, pp. 102–119, 2014.

[2] Olivier Couture et al., "Ultrasound localization microscopy and super-resolution: A state of the art," *IEEE TUFFC*, vol. 65, no. 8, pp. 1304–1320, 2018.

[3] Oren Solomon et al., "Deep unfolded robust pca with application to clutter suppression in ultrasound," *IEEE TMI*, vol. 39, no. 4, pp. 1051–1063, 2020.

[4] Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Variational formulation of unsupervised deep learning for ultrasound image artifact removal," *IEEE TUFFC*, vol. 68, no. 6, pp. 2086–2100, 2021.

[5] Xi Chen et al., "Localization free super-resolution microbubble velocimetry using a long short-term memory neural network," *IEEE TMI*, vol. 42, no. 8, pp. 2374–2385, 2023.

[6] Xin Liu et al., "Deep learning for ultrasound localization microscopy," *IEEE TMI*, vol. 39, no. 10, pp. 3064–3078, 2020.

[7] Baptiste Heiles, Arthur Chavignon, Vincent Hingot, Pauline Lopez, Eliott Teston, and Olivier Couture, "Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy," *Nat. Biomed. Eng.*, vol. 6, pp. 605–616, 2022.

[8] Sungjun Lim et al., "Cyclegan with a blur kernel for deconvolution microscopy: Optimal transport geometry," *IEEE TCI*, vol. 6, pp. 1127–1138, 2020.

[9] Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville, "Improved training of wasserstein gans," in *Neural Information Processing Systems*, 2017, p. 5769–5779.

[10] Wenting Gu et al., "Ultrafast ultrasound localization microscopy by conditional generative adversarial network," *IEEE TUFFC*, vol. 70, no. 1, pp. 25–40, 2023.

[11] Chavignon Arthur, Baptiste Heiles, Hingot Vincent, Lopez Pauline, Eliott Teston, and Couture Olivier, "OPULM PALA," 2020.