Preparation of Papers for IEEE TRANSACTIONS ON MEDICAL IMAGING

First A. Author, *Fellow, IEEE*, Second B. Author,and Third C. Author, Jr., *Member, IEEE*

***Abstract*— High-resolution ultrasound imaging is critical in clinical diagnosis, enabling early detection of abnormalities and precise assessment of anatomical structures. While super-resolution techniques have been widely explored in medical imaging, most existing approaches are restricted to fixed or integer scaling factors. Arbitrary-scale super-resolution, especially for ultrasound images, remains largely unaddressed. This study presents a novel pipeline integrating a lightweight model, EliteNet, with architectural and training modifications to support arbitrary and asymmetric scaling for ultrasound images. A resize layer is introduced at the head of the network to accept user-defined scaling factors, and a two-step training strategy is employed to enhance output quality. In the first step, the model is trained using a hybrid loss combining Structural Similarity Index (SSIM) and Frequency Domain Loss (FDL). In the second step, only the final layer is updated using SSIM and L1 loss, preserving learned features while eliminating artifacts. A dedicated dataset was collected and augmented using flips and reflective padding to ensure structural consistency. Low-resolution images were synthesized using both symmetric and asymmetric scale factors. Our approach yields visually superior results and demonstrates better generalization across arbitrary scales. Quantitatively, it achieves a PSNR of 22.8018 and SSIM of 0.5947, outperforming existing baselines such as ArbRCAN, SRDNet, RDUNet, and ABPN. Extensive ablation studies validate the effectiveness of the loss configuration and training strategy. This work lays foundational groundwork for adaptive, high-quality ultrasound imaging and opens opportunities for real-time, resource-efficient diagnostic applications.**

***Index Terms*— Enter about five key words or phrases in alphabetical order, separated by commas.**

I. Introduction

In medical imaging, the demand for high-resolution images is paramount. Enhanced resolution improves diagnostic accuracy, aids in surgical planning, and facilitates research in disease understanding. Ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) rely heavily on image quality to extract meaningful clinical insights. Among these, US imaging is widely favoured for its real-time capabilities, cost-effectiveness, and non-invasiveness. However, US inherently suffers from resolution limitations due to hardware constraints, noise, and attenuation, which can obscure critical details and impact diagnostic reliability. Super-resolution (SR) techniques have emerged as a powerful tool to address these resolution limitations. Traditional SR methods [1] aim to upscale low-resolution images by fixed scaling factors such as 2x, 3x, or 4x, enhancing visual and diagnostic quality. While effective, these methods often fail to address real-world medical applications' unique and diverse resolution needs. For instance, US imaging in tumor characterization, vascular studies, or fetal assessments may require tailored resolution adjustments to visualize structures of interest optimally. Fixed scaling factors are insufficient to meet such varied demands. Arbitrary Scale Super-Resolution (ASR) [2] offers a transformative approach, enabling image enhancement at any desired scale, including fractional scales such as 1.2x, 1.4x, and 2.3x. This flexibility is particularly valuable in US imaging, where clinicians often require customized resolution enhancements for specific diagnostic tasks. For example, a fractional scaling factor might be essential to highlight delicate structures in vascular imaging or detect subtle abnormalities in soft tissues, which might not be feasible with predefined fixed scales. The ability to achieve arbitrary SR ensures that imaging systems adapt dynamically to the clinical context, improving diagnostic precision and patient outcomes. The development of ASR techniques has been closely tied to advancements in deep learning, particularly Convolutional Neural Networks (CNNs) [3]. CNNs have proven exceptionally effective in modelling spatial hierarchies and extracting complex features from medical images. These capabilities make them well-suited for arbitrary SR tasks, where the goal is to upscale images across a continuous range of scaling factors while preserving anatomical fidelity and minimizing artifacts such as blurring or aliasing. This is especially critical in US imaging, as minor distortions can significantly affect diagnostic accuracy. Despite these advancements, arbitrary scale super-resolution in US imaging presents unique challenges. US images are characterized by high variability in resolution and texture due to differences in probe settings, imaging depth, and tissue properties [4]. Designing models that efficiently learn multi-scale representations and interpolate or extrapolate them at arbitrary scales without introducing artifacts is a complex task. Moreover, the computational demands of ASR must be

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail:author@boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

carefully managed to ensure the technique is viable for real-time applications, which are essential for many ultrasound-based procedures. To address these challenges, this work focuses on modifying the EliteNet network, a state-of-the-art EliteNet model, to enhance its performance for US imaging. The modifications aim to improve the network's ability to accurately handle fractional scaling factors, reducing artifacts and preserving fine anatomical details. By leveraging CNN-based architectures, the proposed approach balances computational efficiency with the need for high-quality super-resolution, making it suitable for the unique demands of US imaging. This paper explores the methodologies underpinning ASR, focusing on CNN-based approaches tailored to US imaging. The discussion encompasses theoretical frameworks and practical implementations, with a detailed evaluation of their effectiveness in handling fractional scaling factors. The study also highlights the broader potential of ASR in medical imaging, underscoring its transformative impact on clinical practice. Finally, ASR's challenges and future directions are discussed, focusing on enhancing adaptability and performance in diverse medical imaging applications.

II. Literature Survey

A. Single image super resolution

Recent advancements in single-image super-resolution (SISR) have significantly leveraged convolutional neural networks (CNNs) to enhance image resolution. The pioneering SRCNN model marked a breakthrough by enabling end-to-end learning of resolution mapping, outperforming traditional methods such as sparse-coding-based super-resolution techniques [5]. Deeper CNN architectures, including VGG-inspired models, followed this, which further incorporated residual learning to improve accuracy [6]. The Enhanced Deep Super-Resolution Network (EDSR) was another milestone, optimizing residual networks and eliminating unnecessary modules to achieve high-quality performance, particularly on benchmark datasets and in competitions like NTIRE 2017 [7]. Other notable methods like LapSRN [8] and DBPN [9] utilized progressive and iterative learning techniques, further improving the resolution. Additionally, innovations such as sub-pixel convolution networks [10] and RCAN [11] have optimized computational efficiency in the field, enabling better performance with fewer resources. These advancements in SISR, which focus on improving the resolution of images, are parallel to innovations across various medical imaging modalities. For instance, achieving higher slice resolution while maintaining a high signal-to-noise ratio (SNR) has long been challenging in fMRI. A modified EPI MRI protocol using slice-shifted images has significantly enhanced slice-direction resolution, improving SNR and better detecting small activated areas in fMRI datasets [12]. Similarly, deep learning models have improved resolution in other imaging modalities, including CT, where leveraging repetitive structures in medical images has resulted in superior image quality compared to traditional methods like SRCNN [13]. In US imaging, resolution constraints due to physical limitations have been addressed through CNN-based approaches, such as an unsupervised super-resolution (USSR) framework that enhances resolution without requiring external datasets [14]. Another technique integrates vision-based interpolation with learning-based US image and video enhancement models, improving spatial resolution and enabling real-time predictions [15]. These approaches echo the successes of deep learning models in SISR, underscoring the potential of CNNs to overcome resolution limitations in diverse medical imaging contexts. In US, deep learning methods such as 3DCNNs [16] and Deep-ULM [17] have significantly improved tissue signal suppression and microbubble localization, further emphasizing the power of deep learning in overcoming resolution challenges in real-time medical applications. These advancements, inspired by techniques in SISR, demonstrate that deep learning is poised to redefine the boundaries of resolution, speed, and clinical precision in medical imaging. SISR methods are effective for enhancing image resolution at fixed scaling factors, but they are limited in their ability to scale images beyond predefined sizes. In contrast, arbitrary scale super-resolution (SR) methods offer greater flexibility, allowing images to be scaled to any desired resolution, thereby enhancing precision and versatility in improving resolution.

B. Arbitrary image super resolution

Super-resolution (SR) techniques have significantly advanced image enhancement in various domains, including medical imaging. However, their application to US imaging remains limited due to challenges such as irregular anatomical features, speckle noise, and low spatial resolution. These characteristics necessitate tailored SR methods that handle the non-linear and complex transformations inherent in US imaging. Our focus is on fractional arbitrary scale SR, a critical advancement for US, which has not yet been adequately explored in the literature. In medical imaging, Arbitrary Scale Super-Resolution (ArSSR) has been introduced in [18], employing an implicit neural voxel function for up-sampling MRI images from low-resolution (LR) inputs into high-resolution (HR) outputs . While this method allows arbitrary scaling, it is restricted to integer scaling and does not address the finer adjustments required for fractional scaling. Similarly, [19] presents an SR model that extends multiscale training with deep convolutional neural networks (CNNs) for MRI, providing flexibility for varying scale factors. Insights from SR techniques for natural images can inform approaches to US SR. The Meta-SR [**?**] model introduces a Meta-Upscale module that dynamically adapts filters for arbitrary scale factors, enabling flexible transformations [20]. SRWarp incorporates adaptive warping layers to address spatially varying transformations, enhancing its ability to handle real-world distortions [21]. Lightweight models like those proposed by [22] integrate scale-aware feature adaptation, balancing computational efficiency with performance for arbitrary-scale tasks. Similarly, Overnet and ASDN refine feature extraction while reducing computational demands for arbitrary-scale SR [23], [24]. While these models offer significant advances, they often assume structured image content, limiting their direct application to the irregularities of US images. Techniques such as combining blind deblurring with SR have proven effective in enhancing robustness

against degradations, which could help handle US-specific artifacts like speckle noise [25]. CiaoSR's scale-aware non-local attention mechanism improves feature representation for arbitrary-scale SR tasks [26]. Additionally, implicit neural representations, such as OPE-Upscale, use position encoding for efficient SR, offering potential benefits for US imaging [27]. Hybrid approaches like MambaSR, combining state-space models with Fourier Convolution Blocks, capture spatial and frequency-domain information, which is valuable for US datasets [28]. Fractional scaling is crucial in US imaging, particularly for clinical applications. For instance, a physician may need to enhance an US scan of a tumor to guide a biopsy. Fractional scaling can improve resolution, helping to delineate tumor boundaries more effectively. Methods like AIDN's Conditional Resampling Module (CRM) show potential for adapting fractional scaling to US SR [29], and the A-LIIF model's use of local implicit functions provides a pathway for ultrasound-specific SR improvements [30]. The existing SR techniques offer a foundation, their adaptation to US imaging requires addressing challenges such as irregular structures, speckle noise, and fractional scaling. Our work aims to develop a novel fractional arbitrary scale SR method tailored for US, addressing these limitations and enhancing its clinical utility.

III. Methodology

A. Dataset

1) Data acquisition: In a typical US process, high-frequency sound waves are transmitted into the body, and the reflected echoes are received, converted into electrical signals, and processed to form images on the machine's display. RF data carries valuable information about the intensity and frequency of these echoes and serves as the foundation for generating various imaging modes, including B-mode, Doppler, and elastography. In this study, RF data was acquired using a research-grade US system equipped with 128 channels. For each transmitted pulse at a specific frequency f, 32 channels operated as transmitters and 64 channels as receivers in a sequential manner, repeated four times to capture the complete echo profile. The probe used for image acquisition was the $L11-5V$ linear array transducer, and scanning was performed on the CAE Blue Phantom breast US model, a high-quality medical training phantom designed to replicate the acoustic and physical properties of real human breast tissue. To generate B-mode US images from the RF data, envelope detection and logarithmic transformation were applied. Envelope detection extracts the amplitude of the raw RF signal received by the transducer, while the log transformation compresses its dynamic range. This enhances the visibility of subtle or low-contrast structures, improving the interpretability of the final B-mode image. The overall data acquisition and processing pipeline is illustrated in Figure 1.

2) Data preprocessing: To prepare the acquired US data for practical model training, a structured preprocessing pipeline was developed to enhance data quality, increase variability, and ensure consistency across all samples. The dataset initially comprised 400 US images captured from a CAE breast phantom at a centre frequency of 5 MHz. The original images

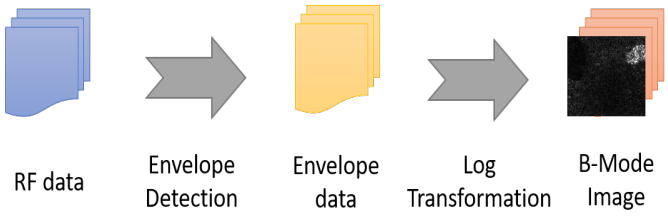
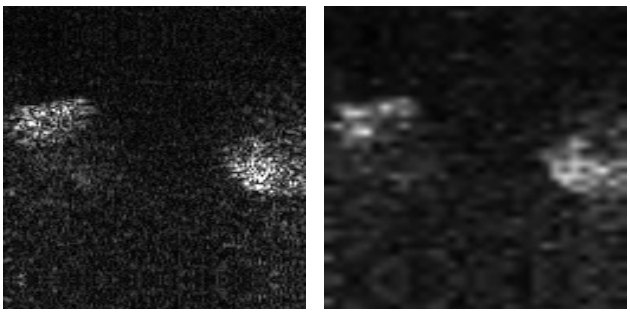


Fig. 1: B-mode image acquisition pipeline

(Figure 2(a)) had a resolution of 128×156 pixels ($w \times h$). To align with the architectural requirements of the model, each image was padded reflectively along the height to reach a uniform size of 128×192. Reflective padding was chosen to preserve edge features and minimize boundary artifacts, which can otherwise distort learning at image borders. Each image was resized down using different scaling factors, resulting in low resolution images. This process allowed the model to learn robust mappings across various resolutions. Finally, a custom dataset class was implemented to handle the multi-scale nature of the data. This ensured that each training batch contained images of different scale factors and their corresponding scaling information, enabling the model to handle variable input conditions effectively and making the training process more flexible and interpretable.



(a) High resolution US image (b) Low-resolution US image

Fig. 2: The figure shows the high and low resolution of the images that we have used in the experiment

B. Multi Objective Loss Design

In medical image super-resolution, preserving high-frequency content is critical, as it directly impacts diagnostic accuracy. To emphasize this, a custom training loss function was adopted, defined as:

$$\mathcal{L}_{\text{train}}(\mathbf{x}, \mathbf{y}) = \alpha \cdot \mathcal{L}_{\text{SSIM}}(\mathbf{x}, \mathbf{y}) + (1 - \alpha) \cdot \mathcal{L}_{\text{FDL}}(\mathbf{x}, \mathbf{y}), \quad (1)$$

Where α is a weighting parameter that balances the SSIM [31] loss and the Frequency Domain Loss (FDL) [32] . The optimal value of α was determined via grid search over the interval $[0, 1]$ with a step size of 0.1, monitored using Weights and Biases to ensure consistent performance across training runs. The SSIM [31] loss function is given by:

$$\mathcal{L}_{\text{SSIM}}(\mathbf{x}, \mathbf{y}) = 1 - \text{SSIM}(\mathbf{x}, \mathbf{y}), \quad (2)$$

Where the Structural Similarity Index (SSIM) is defined as:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\mu_y^2 + \sigma_y^2 + C_2)}, \quad (3)$$

where μ_x is the mean over a window in the image $\mathbf{x}$ and $\sigma_{\mathbf{xy}}$ denotes the covariance between the two images. $\mathcal{L}_{\text{FDL}}(x,y)$ refers to the Frequency Domain Loss, adopted from [32]. FDL defines a perceptual similarity metric that compares two images based on their spatial structures and frequency content. Given two input images x and y, we first extract their deep feature representations using a shared backbone network $\mathcal{F}$. In our implementation, the network chosen was EfficientNet:

$$f_x = \mathcal{F}(x), \quad f_y = \mathcal{F}(y) \quad (4)$$

To analyze the structural and frequency-based differences, we transform these features into the frequency domain via the multidimensional Fast Fourier Transform (FFT):

$$\hat{f}x = \mathcal{F}_{\text{FFT}}(f_x), \quad \hat{f}y = \mathcal{F}_{\text{FFT}}(f_y) \quad (5)$$

and decompose them into magnitude and phase components:

$$M_x = |\hat{f}_x|, \quad \Phi_x = \angle\hat{f}_x, \quad M_y = |\hat{f}_y|, \quad \Phi_y = \angle\hat{f}_y \quad (6)$$

The mean absolute differences between the sorted projections are then computed to quantify dissimilarity in both magnitude and phase:

$$s_{\text{mag}} = \frac{1}{N}\sum_{n=1} |M_x[n] - M_y[n]| \quad (7)$$

$$s_{\text{phase}} = \frac{1}{N}\sum_{n=1}^{N} |\Phi_x[n] - \Phi_y[n]| \quad (8)$$

The final similarity score for each layer i is computed as a weighted sum of the two components:

$$s = s_{\text{mag}} + \lambda \cdot s_{\text{phase}} \quad (9)$$

where λ is a hyperparameter controlling the contribution of phase information, tuned experimentally. The overall perceptual similarity between two inputs is obtained by averaging over all layers:

$$\text{Score}(x,y) = \frac{1}{L}\sum_{i=1}^{L} s^{(i)} \quad (10)$$

where L denotes the total number of feature layers used in the comparison. This combination of SSIM [31] and FDL [32] losses formed the first step of our training process. Although this combination yielded satisfactory results in preserving high-frequency content and structural similarity, it also introduced unexpected white artifacts in the output images. This is likely because the frequency domain constraints do not directly regulate pixel intensity values.

As described extensively in the Experiments Section, to address this challenge, we incorporated a second training step. In the second step, we replaced FDL with L1 loss, maintaining the same overall form of the training loss function:

$$\mathcal{L}\text{train}(x,y) = \alpha \cdot \mathcal{L}\text{SSIM}(x,y) + (1-\alpha) \cdot \mathcal{L}_{\text{L1}}(x,y) \quad (11)$$

where the L1 loss is defined as:

$$\mathcal{L}\text{L1}(I,\hat{I}) = \frac{1}{H \times W \times C}\sum h = 1^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} |I_{h,w,c} - \hat{I}_{h,w,c}| \quad (12)$$

with H, W, and C representing the image's height, width, and number of channels, respectively. Similarly to the first step, the value of α was determined by hyper-parameter tuning. For this stage, the optimal value of α was 0.5. Since the $L1$ loss severely penalizes pixel differences, it effectively reduces white artifacts while preserving the high-frequency details learned in the first step.

IV. Experiments

A. Experimental Setting

1) Dataset Overview: Our experiments utilized a custom US dataset comprising 400 positional images acquired from a CAE breast tumor phantom, including representative tumor samples, cysts, and non-tumor regions. We have applied various data transformations during the preprocessing stage to enhance the model's generalization capability and enable it to learn more robust features. The entire data set was then divided into training (80%), validation (10%), and testing (10%) sets to ensure balanced evaluation at different stages of model development.

2) Implementation details: We utilized the energy-efficient EliteNet model for our experiments due to its ability to deliver high performance while maintaining low computational demands. The model was trained using an NVIDIA Quadro P6000 GPU, which provided the necessary computational power to process US data efficiently. We configured the training with a learning rate of 0.001, which decayed by a factor of 0.8 via ReduceLROnPlateau scheduler over time to ensure gradual convergence. The Adam optimizer was employed to update the network weights, offering adaptive learning rates for different parameters. A batch size of 16 was used and the training was carried out for 500 epochs.The learning rate decayed every 20 epochs if no improvement was seen in the validation loss was seen. These settings were carefully selected to optimize learning and achieve robust US dataset generalization.

We generated low-resolution (LR) training images using symmetric and asymmetric scale factors to develop a model capable of arbitrary scale super-resolution. Symmetric scaling factors ranged from 1.0 to 4.0, with a stride of 0.1, ensuring a dense range of uniform downscaling. Additionally, asymmetric scale factors were introduced by independently varying the horizontal and vertical axes with strides of 0.5, enabling the model to learn non-uniform scaling behaviors. A resize layer was added to the model architecture which resized all LR images to a fixed input size, regardless of their original downscaling factor, thus facilitating arbitrary-scale learning . A similar resizing operation was included to test individual images before passing the image through the model to maintain consistency in input dimensions.

Training was performed in two steps, with each steps involving some key differences. In the initial phase, the loss

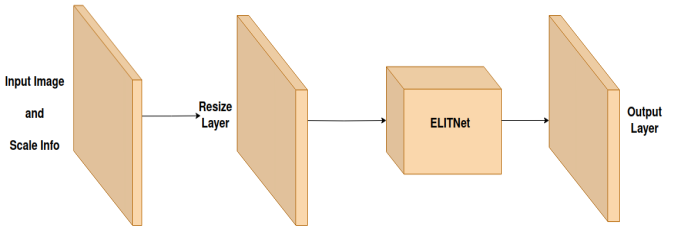


Fig. 3: B-mode image acquisition pipeline

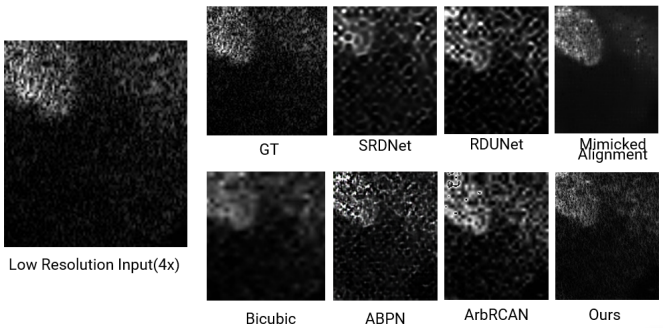


Fig. 4: Comparison of Our Proposed architecture's performance with others

function involved was SSIM [31] + FDL [32] as explained in the Methodology section. Although this combination proved effective in preserving the high frequency components of the image data, this step also introduced random white artifacts in the output image data, for which the second step of training was incorporated, with the loss function being SSIM [31] + L1. An essential aspect of the second training step was freezing all layers except the network's last layer. This allowed the model to retain the high-frequency features learned earlier while adjusting only the final output to eliminate artifacts. Training continued in both training stages until the loss functions exhibited convergence, typically around 500 epochs. Throughout the process, relevant metrics (SSIM [31], PSNR, training loss, and validation loss) were logged to Weights and Biases for effective monitoring.

3) Evaluation Metrics: To evaluate the performance of our trained model in reconstructing high-quality US images, we employed two standard image quality assessment metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [31]. PSNR is a widely used metric that quantifies the difference between predicted and ground truth images. It estimates how much noise or distortion is present in the reconstructed image. Higher PSNR values typically indicate better image quality and closer resemblance to the original image. SSIM [31], on the other hand, measures the perceptual similarity between two images by comparing their structural information, luminance, and contrast. Unlike PSNR, SSIM [31] is more aligned with human visual perception and better indicates how similar the reconstructed image appears to the ground truth.

V. Results

We evaluated the performance of our proposed model by comparing it against several existing super-resolution architectures. The combination of our novel loss functions and multi-step training strategy consistently led to superior results, as reflected in both quantitative metrics and qualitative visual comparisons. Given the limited number of models that support arbitrary-scale super-resolution, we focused our benchmark primarily on the 4× super-resolution task, a standard evaluation scale where improvements are most noticeable and widely reported. The models included in our comparison are: Enhanced Super-Resolution Training via Mimicked Alignment for Real-World Scenes [33], ArbRCAN [34], RDUNet [35], SRDNet [36], and ABPN [37]. Among these, ArbRCAN is one of the few models capable of handling arbitrary scale factors, making it a particularly relevant baseline. While the original ArbRCAN implementation used either L1 or VGG loss during training, we re-trained it using a combination of both losses to ensure a fair comparison, aligning with our objective of preserving both perceptual quality and pixel-level accuracy.

A. Qualitative Results

To ensure a comparison across all super-resolution models, we prepared the training data according to each architecture's expected input format and preprocessing requirements described in their original implementations. This often involved modifying or rebuilding dataset pipelines to match the resolution, normalization, and augmentation strategies used in state-of-the-art configurations. Figure 4 presents the performance comparison for 4× super-resolution, where the improvements were most substantial and clearly highlight the strengths of our approach.

The figure has the qualitative results from each model architecture. One can clearly see the stark difference between the high frequency data preserved by our architecture and that of the others.

B. Quantitative Results

To evaluate the performance of our models, we primarily used two widely accepted image quality metrics: Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). All evaluations were conducted at the 4× super-resolution scale, where the improvements achieved by our proposed architecture were most significant. As shown in Table I, our method consistently outperforms other state-of-the-art models across both metrics.

TABLE I: PSNR and SSIM Comparisons

Model	Scale	PSNR	SSIM
ArbSR	4.00 × 4.00	19.999	0.5867
Mimicked All.	4.00 × 4.00	21.2561	0.3682
SRDNet	4.00 × 4.00	17.495	0.2497
RDUNet	4.00 × 4.00	19.4457	0.3797
ABPN	4.00 × 4.00	18.8270	0.1402
Ours	4.00 × 4.00	**22.8018**	**0.5947**

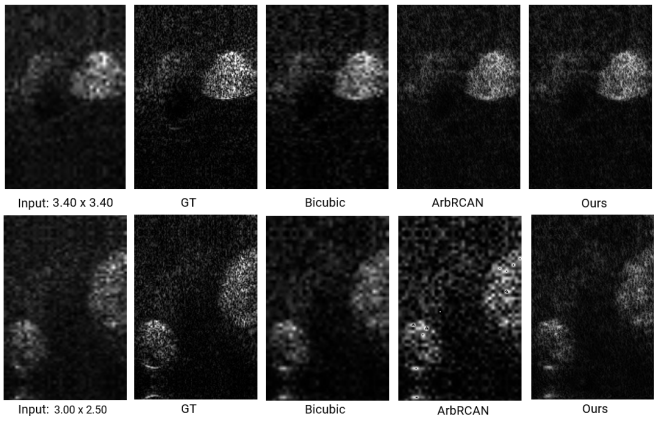


Fig. 5: Comparison with ArbRCAN highlighting the arbitrary scale component.

C. Arbitrary and non-symmetric Scale factors

To demonstrate the flexibility of our model, we evaluated its performance on arbitrary and non-symmetric super-resolution scale factors, an aspect that has received limited attention in prior work. Given the scarcity of architectures explicitly designed for arbitrary scaling, we used **ArbRCAN** [34] as the primary baseline for comparison.

Qualitative examples are provided in Figure 5, while quantitative metrics for select scale factors are shown in Table II. The results indicate that our proposed architecture consistently matches or exceeds the performance of **ArbRCAN**, highlighting its robustness and adaptability across diverse scaling scenarios.

TABLE II: Quantitative Evaluation with ArbRCAN

Scale	Ours		ArbRCAN	
	PSNR	SSIM	PSNR	SSIM
2.50 × 1.50	23.8723	0.6714	23.2296	**0.7501**
3.00 × 2.50	23.1230	0.6146	21.6624	0.6610
3.40 × 3.40	23.0014	0.6059	20.3507	0.6009
2.50 × 1.50	23.8723	0.6714	23.2296	**0.7501**
Overall	**23.7532**	**0.6502**	**22.3573**	**0.6971**

VI. Ablation Study

The primary reason behind training the model in two steps was the preservation of high frequency data and getting rid of any artifacts that would emerge as a result of the first step. Both the steps involved various loss configurations which are detailed in the following sections.

A. First Step

As mentioned in the sections above, the particular loss function combination chosen by us was arrived upon after rigorous experimentation, trial, and errors with several loss functions. This holds true for both the first and the second steps of training. Table III highlights the loss functions that we tested in the first step.

This step was crucial in obtaining a solid frequency-retention loss function, for each step, the loss configurations were used to train the model with the hyper parameters α selected from the range [0.1,0.9] using hyperparameter tuning . All experiments were hosted on W&B with the metrics and the training and validation loss recorded for effective monitoring.

TABLE III: Loss Function Configs for the first step of training.

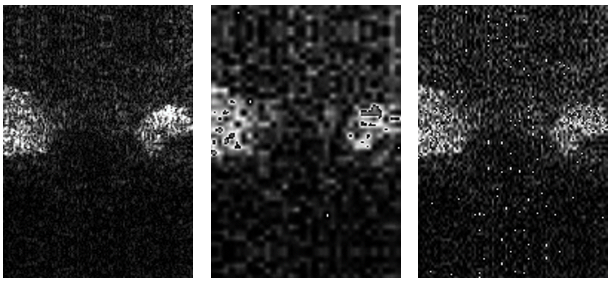
Loss Function	α	Lr.	SSIM	PSNR
SSIM + FFL	0.51	0.001	0.6771	21.457
SSIM + Laplace	0.51	0.001	0.6814	21.4635
SSIM + FDL(EffNet)	0.5	0.001	0.633272	22.971
SSIM + FDL(EffNet)	0.5	1e-5	0.57682	20.4245
SSIM + FDL(VGG)	0.49	0.0004	0.6343	23.50

We have already discussed the SSIM [31] loss function in the methodology section III, here we will briefly dive into the variants used along with it for the first Step. FFL or Focal Frequency Loss [38] is also meant to focus on preserving high frequency components of images and involves fast fourier transforms similar to FDL [32], however one key difference here was the FFL did not involve the extraction of features from the images using a feature extractor(Deep CNNs). Again the performance was good perceptually for smaller scale factors but were not so good for 4x SR..

The second variant tested here along with SSIM [31] was the Laplacian Filter. We used the implementation in the Kornia Computer Vision Library: Laplacian.The Laplacian filter operator smooths the given tensor with a laplacian kernel,it is primarily used for edge detection and image sharpening by highlighting regions of rapid intensity change, something we saw useful in our task.

So the Laplacian loss essentially takes the Super-Resolved output and the ground truth image as inputs and applies the Laplacian filter to both of them. The mean of the absolute difference between the extracted features is returned as the laplacian loss.

The model's performance with this loss configuration was interesting to observing, similar to SSIM [31] + FDL [32] it seemed to preserve the high frequency details for small scales with white artifacts appearing, but the same was not true for higher scaling factors. This was something shared across the other variants as well, even for the first combination - SSIM [31] + FFL [38], SR was satisfactory for small scaling factors, barring appearance of white artifacts in the final output, but for higher scaling factors, the model seemed to hallucinate a lot, as is evident in Fig.6

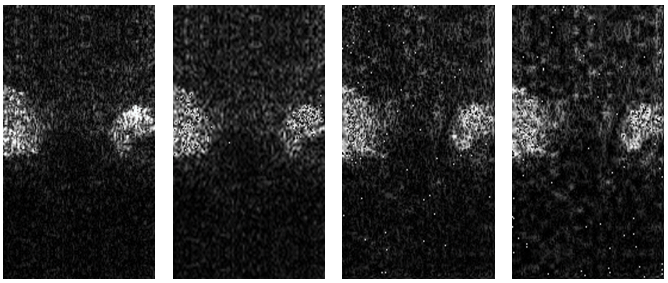


(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5

Fig. 6: Super-resolved outputs using Laplacian loss: (a) Ground Truth (GT), (b) result at scale 4.0×4.0, and (c) result at scale 3.5×1.5. Laplacian loss helps retain edge details across different scales.

SSIM [31] + FDL [32] turned out to be a very promising configuration because it seemed to preserve the overall frequency content of the images both for the small and the large scaling factors. Learning rate and the weight assigned to each loss function in the configuration also played an important role. We have already detailed FDL loss in the methodology section III, however the choice of the feature extractor was also something that we experimented in FDL loss.

Initially the loss function used VGG's feature extraction layers to extract the relevant features. But since the training data varies, the model must also learn the most important features to extract. This meant that a backward pass also involve back propagating through the layers of VGG which made training very time consuming. Hence we switched to EfficientNet as the feature extractor for FDL which cut training time to almost 3/4th of the initial, and allowed us make try out different values of the hyper-parameters. EfficientNet reduced the training without compromising the quality of the final output, as is evident in the comparisons drawn in Fig. 8.



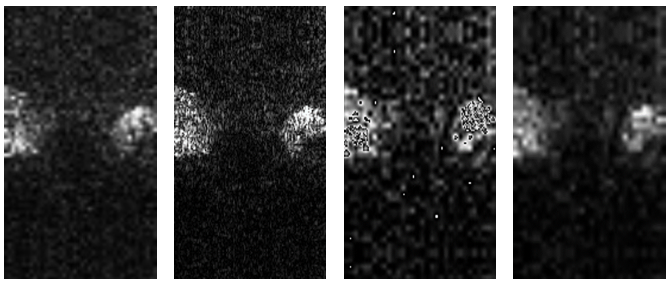
(a) GT (b) EffiNet (c) EffiNet+lr(1e-5) (d) VGG

Fig. 7: Visual comparison of super-resolution outputs using (a) Ground Truth (GT), (b) EffiNet, (c) EffiNet with learning rate $1e{-}5$, and (d) VGG, evaluated under SSIM + FDL settings.

B. Second Step

From the first step, the loss configuration SSIM [31] + FDL [32] proved to be very promising, but as is evident from the output images, there are a lot of white artifacts, especially over the areas indicative of tumor, which is derogatory for diagnosis purposes.

So the purpose of a second step was to train the model to get rid of the white artifacts while holding on to the high frequency data that it had learned to preserve. Therefore we decided to freeze all but the last convolutional layer of the architecture and train all over again. Instinctively, to test if freezing alone would do the job, we did not change the loss function and retrained. But it ended up degrading the output image.



(a) LR Input (b) GT (c) SR (d) Bicubic

Fig. 8: Visual results of super-resolution using SSIM + FDL (EffNet). (a) Low-resolution input, (b) Ground truth (GT), (c) Reconstructed output from our method, and (d) Bicubic interpolation.

Table IV lists out the various loss configurations that we experimented with for the second step. Since the lr and α were chosen from ranges of values between 0.1 and 0.9, we closely monitored the loss curve and metrics that were logged onto W&B and we trained the models with the configurations that showed consistent decrease in validation loss, and promising metrics.

We have discussed all of the loss functions earlier except one. The first configuration in the table CHAR + FDL where CHAR stands for the Charbonnier loss function [39] [40] [41].Once again, we use the implementation from the Kornia library[1]. And that sepcific implementation essentially computes the L1-L2 loss and is computed as follows:

$$\mathrm{WL}(x, y) = \sqrt{(x - y)^2 + 1} - 1$$

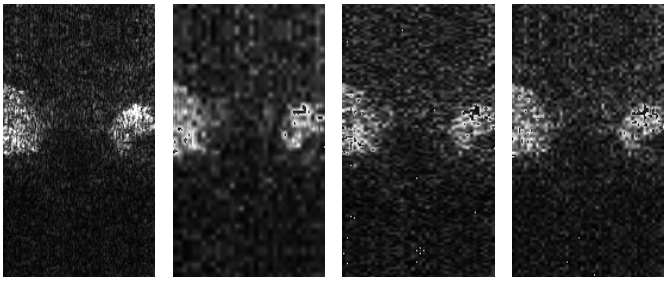
The inspiration for using this loss function came from the results with the L1 loss function, and since the Charbonnier loss incorporates both L1 and L2 loss.

TABLE IV: Loss Configurations that were tried for the second step.

Loss Function	α	**Lr.**	**SSIM**	**PSNR**
CHAR + FDL	0.51	0.0001	0.7032	22.5530
L1 + FDL	0.51	0.001	0.6857	21.3590
L1	0.51	1e-5	0.6168	21.0691
MSE + FFL	0.6	1e-5	0.6290	21.2985
SSIM + FFL	0.51	0.001	0.6771	21.4571
SSIM + FDL(EffNet)	0.57	0.001	0.6778	18.6688
SSIM + L1	0.57	0.001	0.6771	21.4571
SSIM + Laplace	0.51	0.001	0.6393	21.2660
SSIM + Laplace	0.48	0.001	0.6112	19.8712

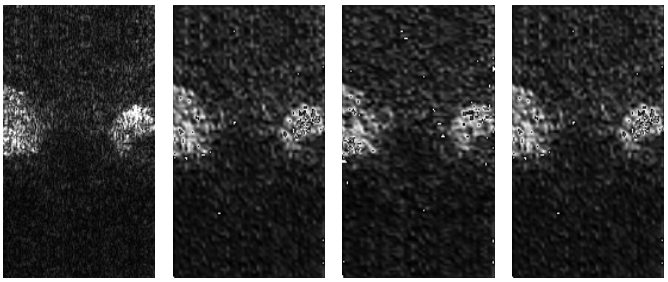
[1]https://kornia.readthedocs.io/en/latest/losses.html#kornia.losses.CharbonnierLoss

But despite decent metrics with Charbonnier loss, the output for higher scaling factors was not perceptually accurate, and the model hallucinated a lot of details, as depicted in Fig.9. All the experiments were run on NVIDIA Quadro P600 GPU, and here are a few Super Resolved Images for each loss combination for different scaling factors:



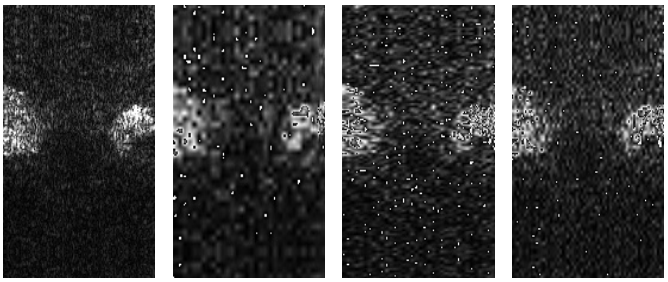
(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 9: Super-resolution results using EffNet with Charbonnier + FDL loss for different scaling factors: (a) Ground Truth, (b) 4.0×4.0, (c) 3.5×1.5, (d) 2.5×2.0.



(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 10: Super-resolution with L1 + FDL: (a) GT, (b) 4.0×4.0, (c) 3.5×1.5, (d) 2.5×2.0.

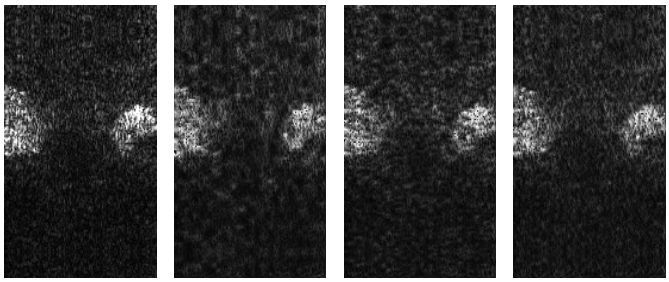


(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 11: Super-resolution results with L1 loss for different arbitrary scale factors: (a) Ground Truth (GT), (b) 4.0×4.0, (c) 3.5×1.5, and (d) 2.5×2.0.

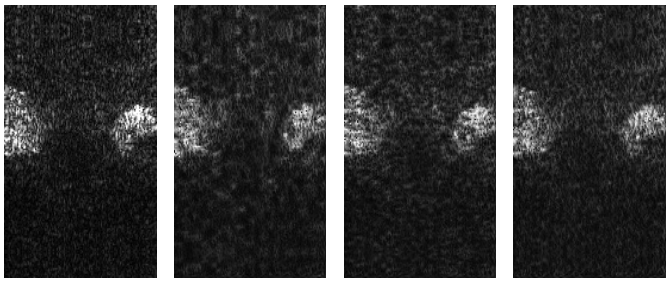
VII. Conclusion

In this study, we leveraged EliteNet, a lightweight super-resolution architecture, to address the challenge of arbitrary and non-uniform scaling in US imaging, a domain with limited prior work. Our contributions include acquiring and preparing



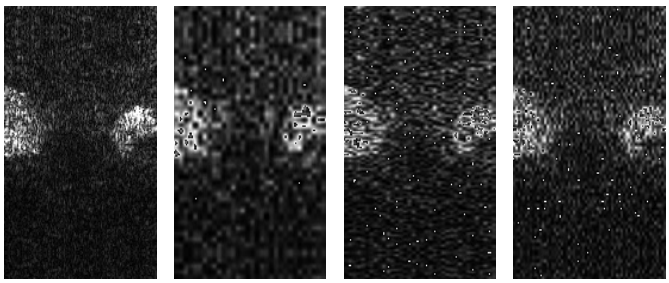
(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 12: Super-resolution results using MSE + FFL loss. (a) Ground truth (GT) image; (b–d) Reconstructed images at different input resolutions: (b) 4.0×4.0, (c) 3.5×1.5, and (d) 2.5×2.0.



(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

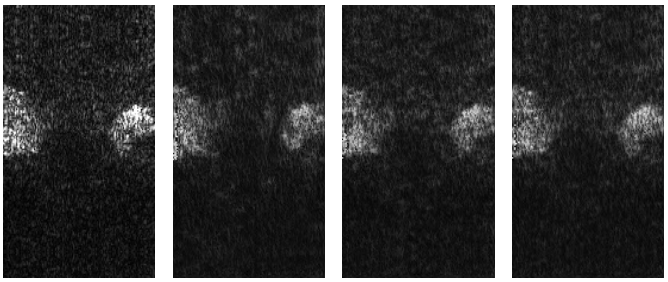
Fig. 13: Super-resolution outputs using Mean Squared Error (MSE) loss. (a) Ground truth (GT); (b–d) Reconstructed images at different low-resolution input scales: (b) 4×4, (c) 3.5×1.5, and (d) 2.5×2.0.



(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

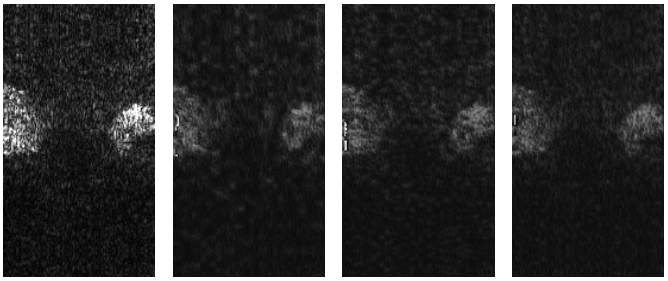
Fig. 14: Qualitative comparison of super-resolved ultrasound images using SSIM + FFL-based enhancement under different scaling conditions. (a) Ground Truth (GT) image; (b) Super-resolved result at a scaling factor of 4.0×4.0; (c) at 3.5×1.5; and (d) at 2.5×2.0.

a specialized US dataset tailored for arbitrary scale super-resolution and designing a novel hybrid loss function that combines L1, SSIM, VGG perceptual loss, and frequency domain loss to improve reconstruction quality. This loss effectively balances pixel accuracy and perceptual fidelity, enhancing performance in PSNR, SSIM, and visual results across various scale factors. Quantitative and qualitative comparisons with baseline models such as ArbRCAN, SRDNet, RDUNet, and ABPN demonstrate that our approach achieves



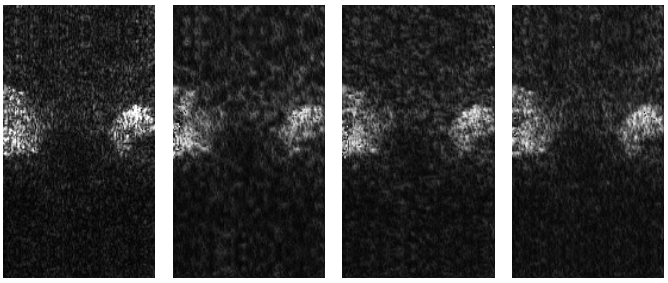
(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 15: Visual comparison of super-resolved ultrasound images reconstructed using different spatial resolution scales with the SSIM + L1 loss function. (a) Ground Truth (GT) image; (b) Super-resolved output at scale 4.0×4.0; (c) Super-resolved output at scale 3.5×1.5; (d) Super-resolved output at scale 2.5×2.0.



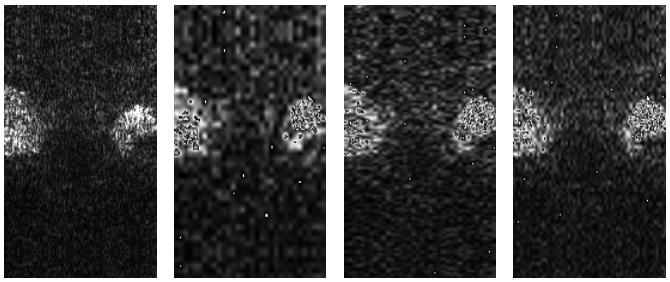
(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 16: Qualitative comparison of super-resolved ultrasound images generated using SSIM + Laplace loss with $\alpha = 0.48$. The subfigures show: (a) Ground Truth (GT), (b) super-resolved output at scaling factor 4.0×4.0, (c) super-resolved output at 3.5×1.5, and (d) super-resolved output at 2.5×2.0



(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 17: Ultrasound image super-resolution results using the combined Structural Similarity Index Measure (SSIM) and Laplace regularization with weighting factor $\alpha = 0.51$. (a) Ground Truth (GT) image representing the original high-resolution tissue scan. (b) Super-resolved image reconstructed with a kernel size of 4×4. (c) Super-resolved image using a kernel size of 3.5×1.5. (d) Super-resolved image generated with a kernel size of 2.5×2.0.



(a) GT (b) 4.0 x 4.0 (c) 3.5 x 1.5 (d) 2.5 x 2.0

Fig. 18: Visual comparison of super-resolved ultrasound images using the SSIM + FDL (EffNet) method. (a) Ground truth (GT) image for reference. (b)–(d) Reconstructed outputs at various super-resolution scaling factors: (b) 4×4, (c) 3.5×1.5, and (d) 2.5×2.0.

superior accuracy while maintaining computational efficiency. The lightweight nature of EliteNet enables practical deployment in low-resource settings and real-time applications. This work provides a comprehensive solution for arbitrary-scale super-resolution of US images by combining a carefully curated dataset, an efficient model architecture, and a powerful learning objective. Future work will focus on automatic scale estimation, modelling temporal consistency for US video, and clinical validation in real diagnostic workflows.

REFERENCES

[1] Karansingh Chauhan, Shail Nimish Patel, Malaram Kumhar, Jitendra Bhatia, Sudeep Tanwar, Innocent Ewean Davidson, Thokozile F. Mazibuko, and Ravi Sharma. Deep learning-based single-image super-resolution: A comprehensive review. *IEEE Access*, 11:21811–21830, 2023.

[2] Hongying Liu, Zekun Li, Fanhua Shang, Yuanyuan Liu, Liang Wan, Wei Feng, and Radu Timofte. Arbitrary-scale super-resolution via deep learning: A comprehensive survey. *Information Fusion*, 102:102015, 2024.

[3] Jianxin Wu. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23):495, 2017.

[4] Johan Thijssen. Spectroscopy and image texture analysis. *Ultrasound in medicine & biology*, 26 Suppl 1:S41–4, 06 2000.

[5] Xiancai Ji, Yao Lu, and Li Guo. Image super-resolution with deep convolutional neural network. In *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, pages 626–630, 2016.

[6] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.

[7] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[8] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.

[9] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018.

[10] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[11] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[12] Ronald R. Peeters, Pierre Kornprobst, Mila Nikolova, Stefan Sunaert, Thierry Vieville, Grégoire Malandain, Rachid Deriche, Olivier Faugeras, Michael Ng, and Paul Van Hecke. The use of super-resolution techniques to reduce slice thickness in functional mri. *International Journal of Imaging Systems and Technology*, 14(3):131–138.

[13] Yunxing Gao, Hengjian Li, Jiwen Dong, and Guang Feng. A deep convolutional network for medical image super-resolution. In *2017 Chinese Automation Congress (CAC)*, pages 5310–5315, 2017.

[14] Jingfeng Lu and Wanyu Liu. Unsupervised super-resolution framework for medical ultrasound images using dilated convolutional neural networks. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 739–744, 2018.

[15] Simone Cammarasana, Paolo Nicolardi, and Giuseppe Patane. Super-resolution of 2d ultrasound images and videos. *Medical & biological engineering & computing*, 61, 05 2023.

[16] Katherine Brown and Kenneth Hoyt. Deep learning in spatiotemporal filtering for super-resolution ultrasound imaging. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 1114–1117, 2019.

[17] Ruud J. G. van Sloun, Oren Solomon, Matthew Bruce, Zin Z. Khaing, Hessel Wijkstra, Yonina C. Eldar, and Massimo Mischi. Super-resolution ultrasound localization microscopy through deep learning. *IEEE Transactions on Medical Imaging*, 40(3):829–839, 2021.

[18] Qing Wu, Yuwei Li, Yawen Sun, Yan Zhou, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. An arbitrary scale super-resolution approach for 3d mr images via implicit neural representation. *IEEE Journal of Biomedical and Health Informatics*, 27(2):1004–1015, 2023.

[19] Chi-Hieu Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, and François Rousseau. Multiscale brain mri super-resolution using deep 3d convolutional networks. *Computerized Medical Imaging and Graphics*, 77:101647, 2019.

[20] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-sr: A magnification-arbitrary network for super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1575–1584, 2019.

[21] Sanghyun Son and Kyoung Mu Lee. Srwarp: Generalized image super-resolution under arbitrary transformation, 04 2021.

[22] Longguang Wang, Yingqian Wang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo. Learning a single network for scale-arbitrary super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4801–4810, 2021.

[23] Parichehr Behjati, Pau Rodriguez, Armin Mehri, Isabelle Hupont, Carles Fernandez Tena, and Jordi Gonzalez. Overnet: Lightweight multi-scale super-resolution with overscaling network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2694–2703, 2021.

[24] Jialiang Shen, Yucheng Wang, and Jian Zhang. Asdn: A deep convolutional network for arbitrary scale image super-resolution. *Mobile Networks and Applications*, 26(1):13–26, 2021.

[25] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep Plug-And-Play Super-Resolution for Arbitrary Blur Kernels . In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1671–1681, Los Alamitos, CA, USA, Jun 2019. IEEE Computer Society.

[26] Jiezhang Cao, Qin Wang, Yongqin Xian, Yawei Li, Bingbing Ni, Zhiming Pi, Kai Zhang, Yulun Zhang, Radu Timofte, and Luc Gool. Ciaosr: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution, 12 2022.

[27] Gaochao Song, Qian Sun, Luo Zhang, Ran Su, Jianfeng Shi, and Ying He. Ope-sr: Orthogonal position encoding for designing a parameter-free upsampling module in arbitrary-scale image super-resolution. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10009–10020, 2023.

[28] Jin Yan, Zongren Chen, Zhiyuan Pei, Xiaoping Lu, and Hua Zheng. Mambasr: Arbitrary-scale super-resolution integrating mamba with fast fourier convolution blocks. *Mathematics*, 12(15), 2024.

[29] Jinbo Xing, Wenbo Hu, Menghan Xia, and Tien-Tsin Wong. Scale-arbitrary invertible image downscaling. *IEEE Transactions on Image Processing*, 32:4259–4274, 2023.

[30] Hongwei Li, Tao Dai, Yiming Li, Xueyi Zou, and Shu-Tao Xia. Adaptive local implicit image function for arbitrary-scale super-resolution. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4033–4037. IEEE, 2022.

[31] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[32] Zhangkai Ni, Juncheng Wu, Zian Wang, Wenhan Yang, Hanli Wang, and Lin Ma. Misalignment-robust frequency distribution loss for image transformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2910–2919, 2024.

[33] Omar Elezabi, Zongwei Wu, and Radu Timofte. Enhanced super-resolution training via mimicked alignment for real-world scenes. In *Proceedings of the Asian Conference on Computer Vision*, pages 4122–4140, 2024.

[34] Longguang Wang, Yingqian Wang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo. Learning a single network for scale-arbitrary super-resolution. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4781–4790, 2021.

[35] Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E Alarcón. A residual dense u-net neural network for image denoising. *IEEE Access*, 9:31742–31754, 2021.

[36] Tingting Liu, Yuan Liu, Chuncheng Zhang, Liyin Yuan, Xiubao Sui, and Qian Chen. Hyperspectral image super-resolution via dual-domain network based on hybrid convolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–18, 2024.

[37] Zhi-Song Liu, Li-Wen Wang, Chu-Tak Li, Wan-Chi Siu, and Yui-Lam Chan. Image super-resolution via attention based back projection networks. In *2019 IEEE/CVF international conference on computer vision workshop (ICCVW)*, pages 3517–3525. IEEE, 2019.

[38] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13919–13929, 2021.

[39] Jonathan Barron. A general and adaptive robust loss function. pages 4326–4334, 06 2019.

[40] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, volume 2, pages 168–172 vol.2, 1994.

[41] Zhengyou Zhang. Parameter estimation techniques: a tutorial with application to conic fitting. *Image and Vision Computing*, 15(1):59–76, 1997.

But despite decent metrics with Charbonnier loss, the output for higher scaling factors was not perceptually accurate, and the model hallucinated a lot of details, as depicted in Fig.9. All the experiments were run on NVIDIA Quadro P600 GPU, and here are a few Super Resolved Images for each loss combination for different scaling factors:

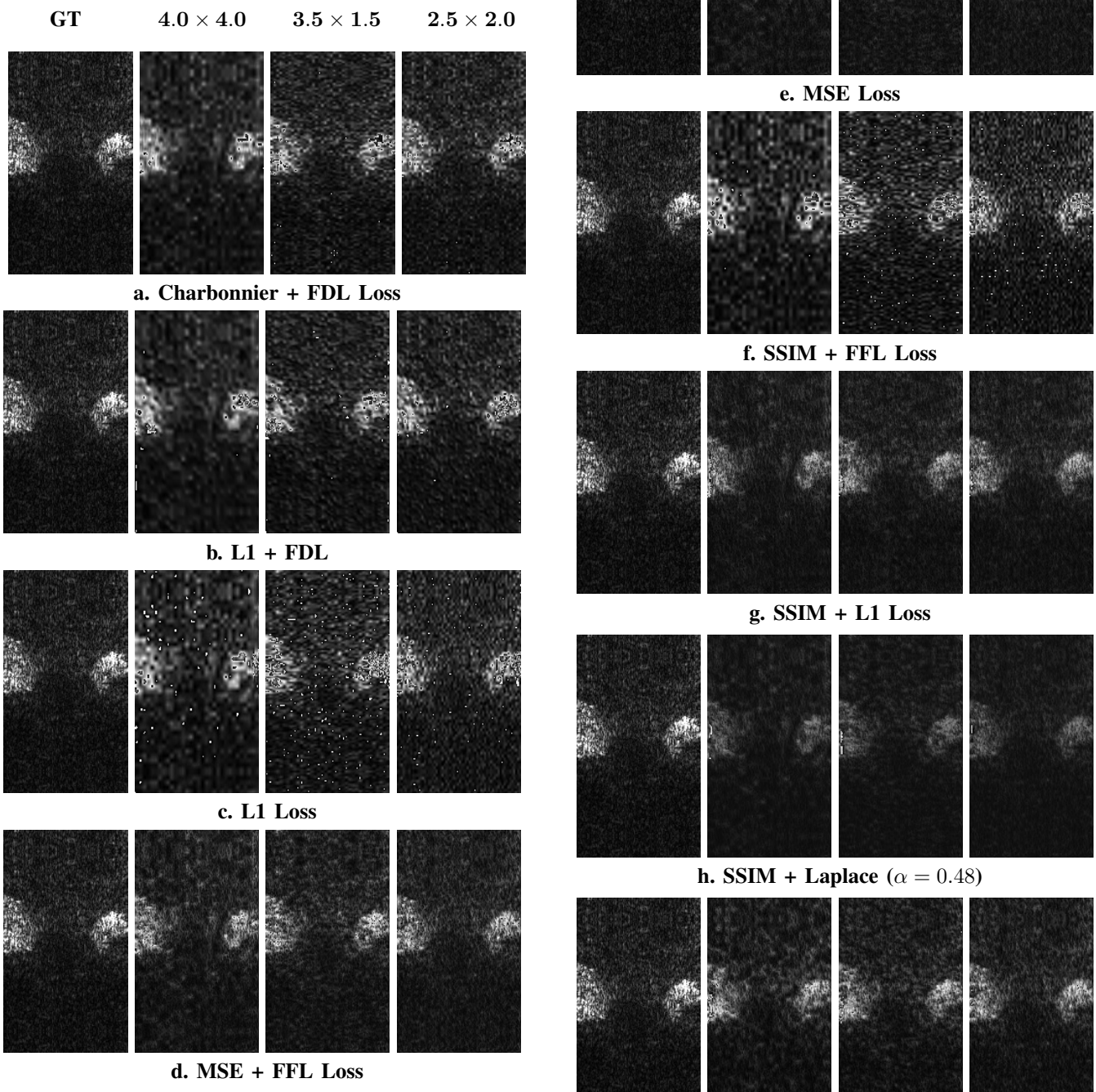


Fig. 19: Qualitative comparison of super-resolved ultrasound images generated using various loss configurations (a-i). The subfigures show: (1) Ground Truth (GT), (2) super-resolved output at scaling factor 4.0×4.0, (3) super-resolved output at 3.5×1.5, and (4) super-resolved output at 2.5×2.0