HDL-SAM: A Hybrid Deep Learning Framework for High-Resolution Imaging in Scanning Acoustic Microscopy

Akshit Sharma[1] Ayush Somani[2] Pragyan Banerjee[3] Frank Melandsø[4] Anowarul Habib[4]

[1]Dept. of Computer Science & Engineering, Indian Institute of Technology Guwahati, India
[2] Dept. of Computer Science, UiT The Arctic University of Norway
[3]Dept. of Mathematics, Indian Institute of Technology Guwahati, India
[4] Dept. of Physics and Technology, UiT The Arctic University of Norway
[1,3]{firstname.lastname}@iitg.in [2,4]{firstname.lastname}@uit.no

Abstract

Scanning acoustic microscopy (SAM) is a cutting-edge label-free imaging technique that allows viewing of both surface and internal structures in a variety of samples, including industrial and biological. Several factors influence the acoustic image resolution, including the signal-to-noise ratio, scanning step size, and the transducer frequency. Our proposed network involves a combination of SwinIR and the hypergraph image inpainting technique specifically adapted to improve the resolution for SAM images. The method aims to fill in missing information, significantly enhancing the resolution of the acquired images. We assessed the effectiveness of our approach against the standalone application of hypergraphs and SwinIR on the dataset, targeting a notable fourfold increase in resolution. The results indicate that the proposed method achieves superior performance, marked by an average structural similarity index measure (SSIM) of 0.92*, a peak signal-to-noise ratio (PSNR) of* 31.60*, and a* $4\times$ *enhancement in image resolution over the raw SAM image. This integration of SwinIR and hypergraphs modules proves indispensable for the precise interpretation of low-resolution acoustic imaging data, allowing the development of reliable tools for image restoration. This improves the fidelity and quality of SAM imaging for both research and practical use.*

1. Introduction

Computer-aided diagnosis combines disciplines such as image processing, machine learning, computer vision, mathematics, physics, and statistics to develop computerized tools for decision-making in fields ranging from material science to biological imaging [7, 40]. Within computer vision, an area of marked interest is high-resolution image inpainting, which involves discretely reconstructing or repairing images [15]. This process is invaluable for repairing damaged regions, removing unwanted features, and seamlessly filling in missing parts of an image. Patch filling finds various applications, including object removal, high-resolution imaging, and image denoising.

The primary challenge in refining high-resolution image inpainting is to seamlessly blend global semantic context and local texture details consistent with the background [45]. Advances in deep learning have risen to this challenge by predicting missing information using existing image data, outperforming traditional inpainting techniques, and improving output quality. Despite its established effectiveness in biomedical research, its application to SAM imaging remains difficult due to the limited availability of training data. SAM images are complex, with varied degrees of contrast and noise, along with the uneven distribution of missing data. They lack a holistic understanding of the global context or semantics of the image, leading to results that are often implausible. Some acoustic microscopy imaging applications may also demand real-time non-invasive processing. Nevertheless, high-resolution acoustic images are critical to biomedical and materials research. They allow for investigating, measuring, and determining the mechanical or biomechanical properties of various samples [18, 19].

1.1. Scanning acoustic microscopy

Scanning acoustic microscopy (SAM) is a label-free imaging technique widely used in biomedical imaging, non-destructive testing, and material research fields. It allows for high-resolution visualization of both surface and subsurface structures with precision. In addition to its inspection capabilities, SAM provides extensive and accurate quantitative data about the inspected objects. SAM possesses a range of capabilities, including non-invasive microstructural characterization of materials and monitoring the structural health of composite structures [17, 18, 20]. It also detects surface defects in polymer circuits and analyzes anisotropic phonon propagation [4, 21]. SAM is of significant importance in the highly competitive microelectronics and semiconductor industries [38].

The clarity of images in Scanning Acoustic Microscopy (SAM) relies on various factors such as excitation signal frequency, signal-to-noise ratio, and pixel size. At a given frequency, SAM image quality hinges on the pixel size or scanning steps in both horizontal (x) and vertical (y) directions, alongside the acoustic beam's size. Lower-resolution images require fewer scanning points, hence speeding up the process, whereas higher resolution demands more points, elongating data acquisition. In imaging biological specimens, data acquisition is vital, favoring high resolution with finer step sizes. However, larger scanning steps can compromise image quality due to information loss. Despite the time-consuming pixel-by-pixel scanning method in SAM, achieving high-resolution images entails employing a high-frequency transducer and finer step sizes. Conversely, low-resolution ultrasound images cover fewer points, reducing scanning time. The utilization of a super-resolution model simplifies the acquisition of high-resolution ultrasound images effortlessly.

Deep learning-based algorithms are adept at identifying patterns in naturally occurring images, making deep neural networks a common choice for computer vision tasks due to their ability to deliver superior results over a shallow network model [34]. However, the absence of clear patterns in ultrasound imaging data suggests that blindly employing very deep neural networks for resolution upscaling may not align with our objectives. With the advent of new technologies that facilitate the generation of high-quality images, traditional methods of image generation are becoming outdated. Consequently, researchers are making efforts to improve high-resolution imaging techniques [2, 33].

In this paper, we revisit the high-resolution image inpainting technique, with a special emphasis on its application to acoustic microscopy imaging. To our knowledge, this marks the first investigation into the use of a synergistic combination of hypergraphs [37] and high-resolution techniques (SwinIR) [28] for acoustic microscopy. The proposed network is capable of capturing the intrinsic details from a low-resolution acoustic microscopy image, revealing that the hybrid framework not only achieves state-of-the-art (SOTA) performance but significantly improves the SSIM and PSNR scores beyond what their applications can provide. The methodology addresses and potentially overcomes the constraints posed by the step size limitations inherent in high-resolution imaging for live biological samples, where time is a crucial factor. Figure 1 outlines the overall strategy used in this paper.

2. Related Work

Achieving super-resolution in SAM presents a multi-faceted challenge, encompassing issues related to image quality, resolution enhancement, and interpretability. Over the years, researchers have explored diverse methodologies to tackle these challenges, drawing upon advancements in image inpainting and image super-resolution techniques.

Image inpainting methods aim to fill missing or damaged regions in images while preserving global semantic structures and fine details. Generating images that are contextually plausible and realistic involves addressing coherence in global semantic structures and finely detailed texture around missing areas. Image inpainting techniques tackle this challenge through various approaches, most notably through content/texture adaptation methods and learning-based methods.

Content/texture adaptation methods heavily relied on patch-matching algorithms for inpainting [3, 10]. These approaches iteratively fill missing pixels by searching for similar patches from neighboring non-missing pixels. While effective in synthesizing texture-consistent outputs [44], they struggled to produce semantically meaningful content. Some recent advancements, such as exemplar-based methods [12], have shown promise in filling geometric patterns and textures. However, they predominantly rely on low-level features for patch matching and are limited in filling voids with semantically meaningful or novel content.

On the other hand, learning-based techniques, particularly Generative Adversarial Networks (GANs), proved a powerful strategy for image inpainting [5, 23]. GANs consist of a generator trained on a dataset to generate new examples and a discriminator that ensures the generated outputs are plausible within the dataset domain. Recent advances include the integration of contextual attention layers [41] and gated convolutions [42] for improved performance. However, some models struggle with irregular masks, prompting the development of techniques such as partial convolution [29] to handle such cases effectively. Despite their effectiveness, learning-based models often lack interpretability, making inpainting output control challenging.

Parallelly, image super-resolution techniques have emerged as a pivotal area of research, aiming to elevate the visual quality of low-resolution images by reinstating finer details. There has been a significant push over the last decade to improve the performance of deep learning methods. Notably, within biomedical imaging and MRI, the refinement of single-image super-resolution methods has been a focal point, as evidenced by studies such as those by Yuan et al. [43] and Tang et al. [35]. The inception of pioneering architectures like SRCNN by Dong et al. [13] helped numerous studies explore various architectural enhancements to improve performance, including those pro-

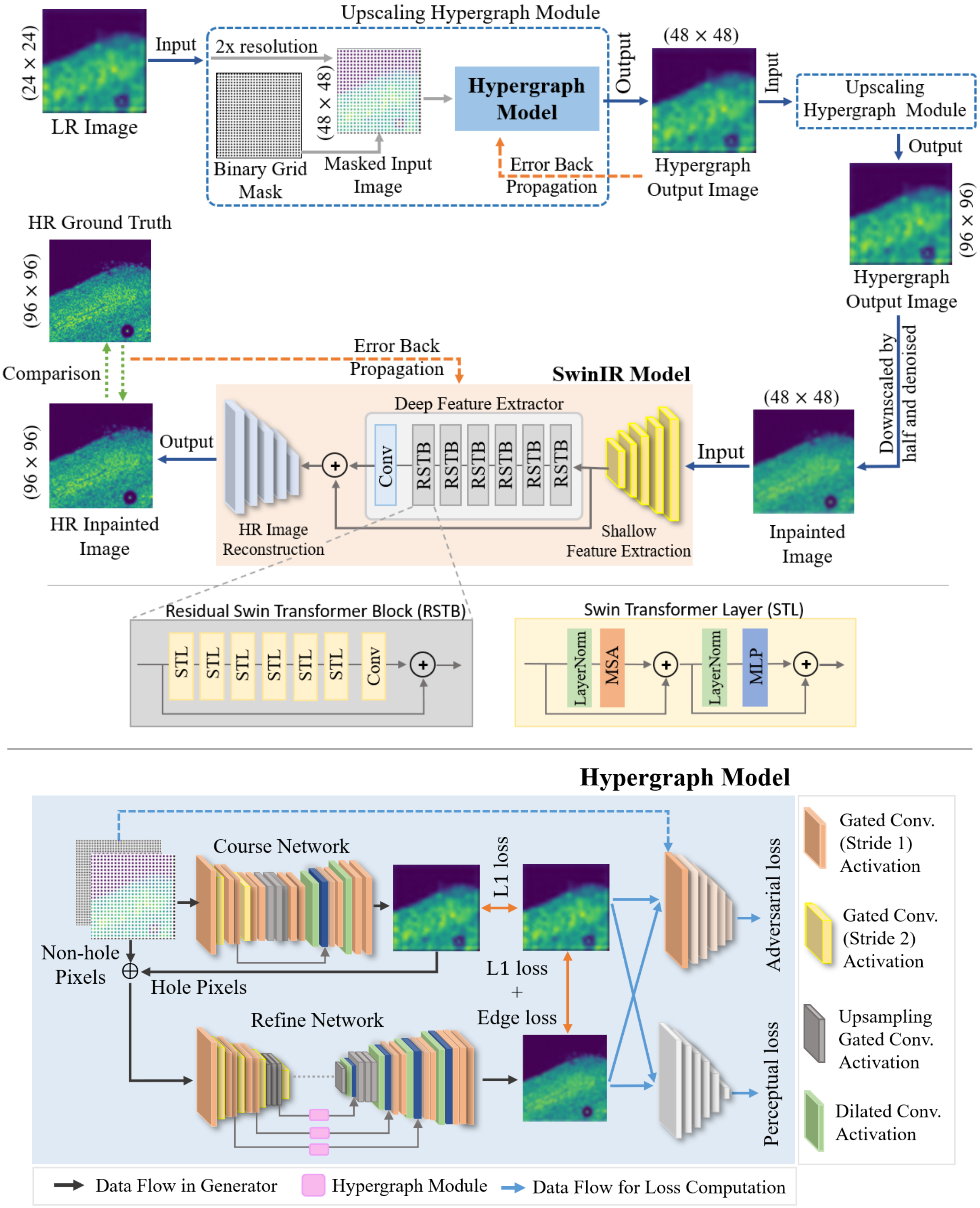

Figure 1. This illustration demonstrates the high-resolution image inpainting strategy for SAM images, employing SwinIR and hypergraph techniques. The alternative hole mask serves as the basis for generating a low-resolution (LR) input image for the model. Subsequently, image inpainting techniques are employed to enhance the resolution, resulting in the creation of a high-resolution (HR) version of the original image.

posed by Dai et al. [11] and Kim et al. [25], all aimed towards improving performance metrics.

Furthermore, the development of autoencoder-based approaches [32] and GANs like EE-SRGAN [30] specifically for biomedical images has enabled more precise diagnostic procedures and refined treatment strategies. More recently, the introduction of Transformer-based models to image restoration tasks such as SwinIR [28], CAT [8], ELAN [47], ART [46], and GRL [27] have shown remarkable efficacy. These models excel by expanding the scope of receptive fields, achieving superior results in image impainting.

In the field of high-resolution imaging in Scanning Acoustic Microscopy (SAM), some studies utilized U-NET-inspired architectures [31]. Whereas the recent advancements have shifted towards transformer-based approaches [2, 36] owing to their capability to capture global interactions and outperform CNNs in various tasks [6, 14]. Transformer networks, inspired by the attention mechanism, have demonstrated remarkable performance in tasks that require an understanding of long-range dependencies and the global context.

We believe, that adopting a hybrid approach that combines image inpainting with super-resolution image restoration techniques specifically for SAM imaging can emerge as an innovative and effective strategy to tackle the complex challenges of SAM imaging. This approach aims to advance SAM imaging resolution using computer-aided image processing, offering a promising pathway to overcome obstacles and unlock new potential in acoustic data interpretation.

3. Proposed Method

In this paper, we introduce HDL-SAM, a novel hybrid model that leverages the combination of low-resolution image inpainting and high-resolution image restoration to achieve a fourfold improvement in image resolution. The method begins with the crucial task of image inpainting, where the reconstruction of missing regions in an image is achieved using a hypergraph-based architecture that is heavily based on the research methodology described in [37]. The method involves the fabrication of a mask to train the dataset, specifically employing a pattern in which, for every three white pixels, a black matrix is introduced. By adding three white pixels between each recorded data point, we effectively expand the lower-resolution image by a factor of four.

Attempting to train and apply the hypergraphs model directly for $4\times$ image inpainting were at par with current standards but failed to give us SOTA performance. Consequently, we adapted our strategy by training the hypergraph from scratch to perform $4\times$ upsampling in two stages (shown in Fig. 1) - each stage involving $2\times$ upsampling. This intermediate $2\times$ upsampling was achieved by inserting a white pixel between each recorded data point, which was then filled in during the image inpainting process. Low-resolution images, being less prone to noise and primarily composed of the scene's major structures, present fewer local orientation singularities that might complicate the filling order. Furthermore, working with a smaller image before inpainting significantly reduces the computational time required to inpaint the full-resolution image. This approach underlines the rationale for prioritizing image inpainting before undertaking high-resolution restoration.

For the image restoration phase, we turn to the SwinIR architecture [28], selected for its efficacy in addressing various forms of image degradation, such as blurring, noise, and distortion, through an inverse process that aims to closely approximate the original, undistorted image. The integration of the SwinIR model as part of HDL-SAM is not just for its advanced capabilities in improving image detail and clarity, but also for its complementarity to our inpainting process, ensuring that the transition from inpainted low-resolution images to high-resolution outputs is seamless and efficient. This holistic approach, combining hypergraph-based inpainting for filling in missing grid mask areas with the SwinIR architecture for overall image restoration and resolution enhancement, sets a new standard in the field by addressing both specific and general aspects of image quality improvement. In the following sections, we outline the core concepts of our research and highlight the innovative aspects of our approach to achieving high resolution.

3.1. Hypergraphs

The hypergraph combines spatial and feature-based clustering strategies to capture both local and global structures in the I'm age. Formally represented as $G = (V, E, W)$, it comprises a set of vertices $V = \{v_1, ..., v_n\}$ and hyperedges $E = \{e_1, ..., e_n\}$, where each hyperedge can connect two or more vertices. The diagonal matrix $W \in \mathbb{R}^{M\times M}$ holds the weight of each hyperedge. Additionally, the structure of the hypergraph G can also be characterized by an incidence matrix $H \in \mathbb{R}^{N\times M}$, where the link is illustrated in Equation 1.

$$h(v, e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{if } v \notin e \end{cases} \tag{1}$$

Simple graphs can be seen as a special case of hypergraphs, with each hyperedge connecting only two nodes. They can easily represent the pairwise data relationships, but it is difficult to represent the spatial features and their relationship in an image. Hence, hypergraphs are used instead of graphs. To transform the spatial features $F^l \in \mathbb{R}^{hw\times c}$ into a graph-like structure by treating each spatial feature as a node with dimension c, and a feature vector, $X^l \in \mathbb{R}^{hw\times c}$.

For the incidence matrix H, instead of using the Euclidean distance between features of images [16, 39], crosscorrelation of the spatial features is used to calculate the contribution of each node to the hyperedge. As a result,

$$H = \Psi(X)\Lambda(X)\Psi(X)^T\Omega(X) \tag{2}$$

where $\Psi(X) \in \mathbb{R}^{N\times C}$ represents the linear embedding of the input features followed by the ReLU activation, and $\hat{C}$ denoting the dimension of the vector of features post-embedding. $\lambda(X) \in \mathbb{R}^{\hat{C}\times\hat{C}}$ is a diagonal matrix that helps to learn a better distance metric among the nodes for the incidence matrix H, and $\Omega(X) \in \mathbb{R}^{N\times M}$ helps to determine the contribution of each node to every hyperedge, where m is the number of hyperedges in the hypergraphs. Implementing $\Psi(X)$ involves a 1×1 convolution on the input features. Meanwhile, $\Lambda(X)$ is implemented by channel-wise global average pooling followed by a 1×1 convolution as stated in [22], and $\Omega(X)$ is implemented using the 7×7 filter. Consequently, we arrive at:

$$H^l = \Psi(X^l)\Lambda(X^l)\Psi(X^l)^T\Lambda(X^l)^T \tag{3}$$

$$\Psi(X^l) = \text{conv}(X^l, W^l_\Psi) \tag{4}$$

$$\Lambda(X^l) = \text{diag}(\text{conv}(\hat{X}^l, W^l_\Lambda)) \tag{5}$$

$$\Omega(X^l) = \text{conv}(X^l, W^l_\Omega) \tag{6}$$

where, $\hat{x}^l \in \mathbb{R}^{1\times 1\times\hat{C}}$ is the feature map produced after global pooling of the input features, and $W^l_\Psi, W^l_\Lambda, W^l_\Omega$ are the learnable parameters for linear embedding. Absolute values are used in the incident matrix to avoid imaginary values in the degree matrices. Hence, the hypergraphs convolution layer on spatial features can be written in Equation 7 as,

$$X^{l+1} = \sigma(\Delta X^l\Theta) \tag{7}$$

where $\Theta \in \mathbb{R}^{C_l\times C_{l+1}}$ is the learnable parameter and σ is the ELU [9] non-linear activation function.

3.2. SwinIR

SwinIR comprises three main modules: shallow feature extraction, deep feature extraction, and high-quality image reconstruction. Initially, the shallow feature extraction module uses a 3×3 convolution layer, $H_{SF}(.)$ to extract stable features F_0 from input image of low quality, I_{LQ}. This process effectively maps the input image to a higher-dimensional feature space. Following this, the deep feature extraction phase uses K residual Swin transformer blocks (RSTB) along with a 3×3 convolutional layer to extract a series of intermediate features from F_1 to F_k, leading to the final deep feature F_{DK}. Each RSTB applies self-attention and feed-forward layers to the image patches, adhering to the core deep learning principles.

For the reconstruction of the high-quality image, I_{RHQ} the model aggregates shallow and deep features, utilizing a function $H_{REC}(.)$ to combine F_0 and F_{DF}. This process aims to capture low-frequency details through shallow features while retrieving high-frequency details lost in the original low-quality image using deep features. The reconstruction module employs sub-pixel convolutional layers for upsampling and utilizes residual learning to reconstruct the residual between the low-quality and the high-quality images.

SwinIR parameters are optimized by minimizing the L_1 pixel loss between the reconstructed high-quality image, I_{RHQ}, and the ground truth high-quality image, I_{HQ}. Each RSTB (shown in Fig. 1) comprises Swin transformer layers (STL) and convolutional layers. The Swin transformer layer differs from the standard transformer layer by incorporating local attention and a shifted window mechanism. It computes $query$, key, and $value$ matrices for a local window feature X, which are then used in the self-attention mechanism. Additional feature transformations are performed using the multi-head self-attention and multi-layer perception (MLP) modules, each with LayerNorm and residual connections.

3.3. Dataset setup and training strategy

In our study, we used SAM with a 50 MHz transducer to generate a collection of 33 high-resolution images, each enhanced to four times their original resolution. The acquisition process involved a step size of 50 μm on both axes, resulting in images of varying sizes and aspect ratios. Each image is cropped into several 96×96 pixels sub-images starting from the image's upper left corner and strides in the G-direction and H-direction. This was done to maintain a uniform size for training while also ensuring that the overall semantics of the images are somewhat preserved.

The dataset comprised 402 such 96×96 images, normalized to a range [0-1] during training. For every high-resolution crop, a corresponding low-resolution (20 MHz transducer) crop was created by masking 3 pixels in a 2×2 window with a stride size of 2, retaining only the upper left pixel in each of these windows. The mask is inspired by the fact that the SAM operates at two different step sizes. The low-resolution images are recorded to have a step size of $200\mu m$, while the higher-resolution images used for model training have a step size of $50\mu m$.

This dataset of 402 images is split into two subsets: 306 training (set A) and 96 testing images (set B), respectively. Set A with 24×24 low-resolution images, is used for training the first model of the pipeline in a two-step iterative process, focusing on hypergraph-based image inpainting. Its performance is assessed using set B. This trained hypergraphs model is executed on set B to generate set C of images as output (consisting of $4\times$ upscaled images, measuring 96×96) in two steps. These images were then downscaled by half using the Lanczos resampling algorithm, and passed through a denoising model to mitigate any artifacts from downscaling, resulting in set D. After this refinement step, a selection of 66 images from set D underwent geometric data augmentation to produce 330 images (48 × 48), serving as the training set for the SwinIR model. The trained SwinIR was subsequently evaluated on the remaining images to achieve high-resolution output.

To demonstrate the effectiveness of combining the two models (see Fig. 1) for high-resolution imaging, each model was trained independently on the entire dataset and on a subset of the pipeline. The hypergraphs model was specifically trained on set A for two-step 4× upscaling in two steps and evaluated on set B. The training was done from scratch, with a batch size of 1 for 200 epochs. We applied transfer learning, using a base model initially trained on the CelebA-HQ dataset. For the SwinIR model, one instance was trained on set D (48 × 48 images) and evaluated on the remaining images. Another instance was directly tasked with 4× upscaling on set A to underline our proposed framework's efficiency compared to using SwinIR or hypergraphs independently. Training for both SwinIR instances involved transfer learning from the model pretrained on DIV2K [1] + Flickr2K [28] dataset with a batch size of 8 over 3500 epochs, using L1 loss with an exponential moving average of 0.96. Adam optimizer was used with a learning rate 2×10^{-4} and no weight decay.

3.4. Experimental setup

The SAM functions in both reflection and transmission modes, with a labeled diagram of the SAM setup for image acquisition shown in Figure 2. The intricacies of these operational modes that can be found in existing literature are beyond the scope of this paper. We focus primarily on the reflection mode for scanning the samples, where acoustic energy is channeled through a coupling medium, typically water, using a concave spherical sapphire lens rod, which facilitates the examination of samples. The process involves generating ultrasound signals that interact with the sample, with the reflections captured and digitized, termed the A-scan, or amplitude scan. By replicating this step across various points on the XY plane, we aggregate these A-scans to produce a detailed two-dimensional C-scan image, offering a nuanced view of the sample's acoustic properties.

Data for this study were collected using a custom-designed SAM system, featuring a high-precision scanning stage from Standa (8MTF-200-Motorized XY Microscope Stage) and operated through a LabVIEW program. This setup, mirroring the one used in previous research by Kumar *et al.* [26] to correct for samples at an angle, incorporates acoustic imaging capabilities using National Instruments' PXIe FPGA modules and FlexRIO hardware, housed within a PXIe chassis (PXIe-1082), and includes an arbitrary waveform generator (AT-1212). The transducers were excited with Mexican hat signals and amplified by an RF (radio frequency) amplifier (AMP018032-T) to strengthen the ultrasonic signals. Reflected acoustic waves from the sample are then amplified through a tailor-made amplifier, further improved by a custom pre-amplifier, and digitized with a high-speed 12-bit (1.6 GS/s) digitizer (NI-5772).

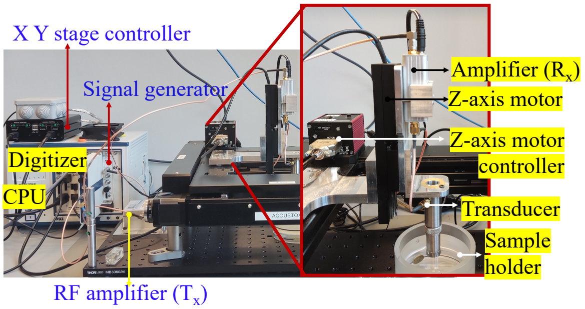

Figure 2. This labeled illustration presents a detailed overview of the SAM system employed for image acquisition, effectively outlining and describing each essential component within the experimental setup

A 50 MHz focused transducer from Olympus with a 12.5 mm focal length and a 6.35 mm aperture, provided the standard for accuracy. This transducer was used to scan both a coin and a biological specimen, concentrating the acoustic energy on the sample's top surface. Scanning was performed in the x and y-directions with $50\mu m$ steps. Low-resolution images were captured using a 20 MHz transducer with a 50 mm focal length. All experiments were carried out in distilled water at a steady room temperature of approximately 22°C. A discarded reindeer antler served as the biological sample for imaging evaluation. Before scanning, the antler was cleaned of moss with lukewarm water and 96% ethanol, then boiled in distilled water at 100°C to remove any residual biological substances. The sample was then placed on the sample holder to dry before scanning.

The performance of HDL-SAM is validated against other models such as SwinIR, hypergraphs, and AOT-GAN for 4× resolution enhancement. This validation is conducted through extensive training and testing on the dataset.

4. Results and Discussion

After the initial training phase, the hypergraphs module was assessed using the CelebA-HQ dataset [24] shown in Figure 3. A PSNR score of 22.94 and an SSIM score of 0.79 were obtained. Following the application of transfer learning techniques, the hypergraphs module underwent further evaluation on set B, as detailed in Section 3.3, achieving a PSNR score of 25.68 and an SSIM score of 0.78.

When SwinIR was specifically trained for 4× upscaling

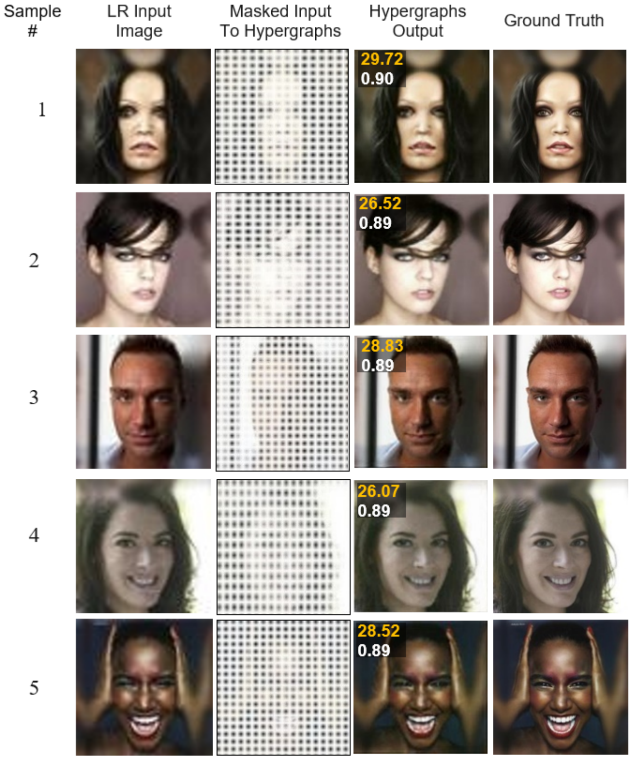

Figure 3. This randomly sampled result illustrates the effectiveness of the hypergraphs model when utilized for enhancing image resolution on the CelebAHQ dataset. Specifically, the model aims to upscale the image resolution by a factor of 4×, demonstrating its capability to generate high-quality images with improved resolution.

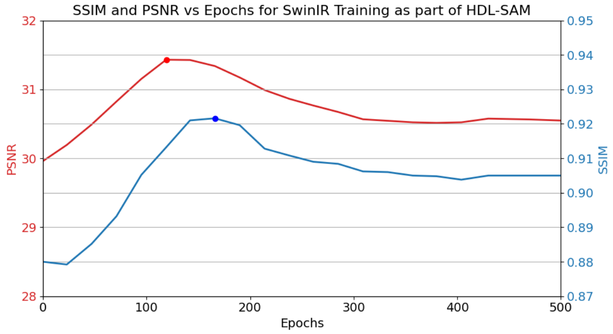

Figure 4. Training progression for SwinIR instance within the HDL-SAM framework, specifically for the task of 2× upscaling. The graph presents the average PSNR (displayed in red) and SSIM (depicted in blue) throughout a training period of 500 epochs.

directly, its performance on Set B obtained a PSNR of 28.15 and an SSIM of 0.84 after 220 epochs. In contrast, HDL-SAM achieved a peak PSNR of 31.17 and an SSIM of 0.92 in just 119 epochs. Across other benchmarked models on the entire testing dataset, AOT-GAN reported a PSNR score of 27.81 and SSIM of 0.75, while Swift-SRGAN showed comparatively lower performance, with an average PSNR of 14.57 and an SSIM of 0.60. Training progress plots for both SwinIR instances are shown in Figures 4 and 5, with a summary of these findings presented in Table 1.

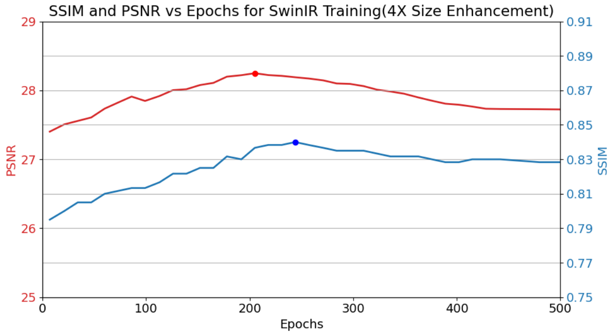

Figure 5. Training progression for SwinIR instance in HDL-SAM, particularly highlighting its effectiveness in achieving a 4× upscaling task. The graph presents the average PSNR (depicted in red) and SSIM (represented in blue color) throughout a training period of 500 epochs.

	Swift-SRGAN	Hypergraphs	SwinIR	AOT-GAN	HDL-SAM
PSNR	14.57	25.68	28.15	27.81	**31.60±1.18**
SSIM	0.60	0.78	0.84	0.75	**0.92±0.01**

Table 1. Average SSIM & PSNR scores of several models including HDL-SAM when applied independently on the testing set.

Using the visual Turing test (VTT), it is observed that the images upscaled by 4× using HDL-SAM significantly outperform the results obtained by other benchmarked models, as shown in Figure 6. The quantitative improvement achieved by HDL-SAM over all other models (averaged) across each randomly sampled test image is presented in the figure, providing a clear basis for comparison.

4.1. Resource-efficient learning through proposed hybrid deep learning approaches

The rapid rise of deep learning capabilities has unsurprisingly resulted in groundbreaking advancements in various domains, including image processing. However, the environmental impact and resource-intensive training of SOTA deep neural networks from scratch have raised serious concerns regarding sustainable AI development. This paper introduces a novel hybrid model that makes an effort to address these concerns by prioritizing resource efficiency and lowering the carbon footprint associated with deep learning model training.

The approach values the use of existing SOTA models

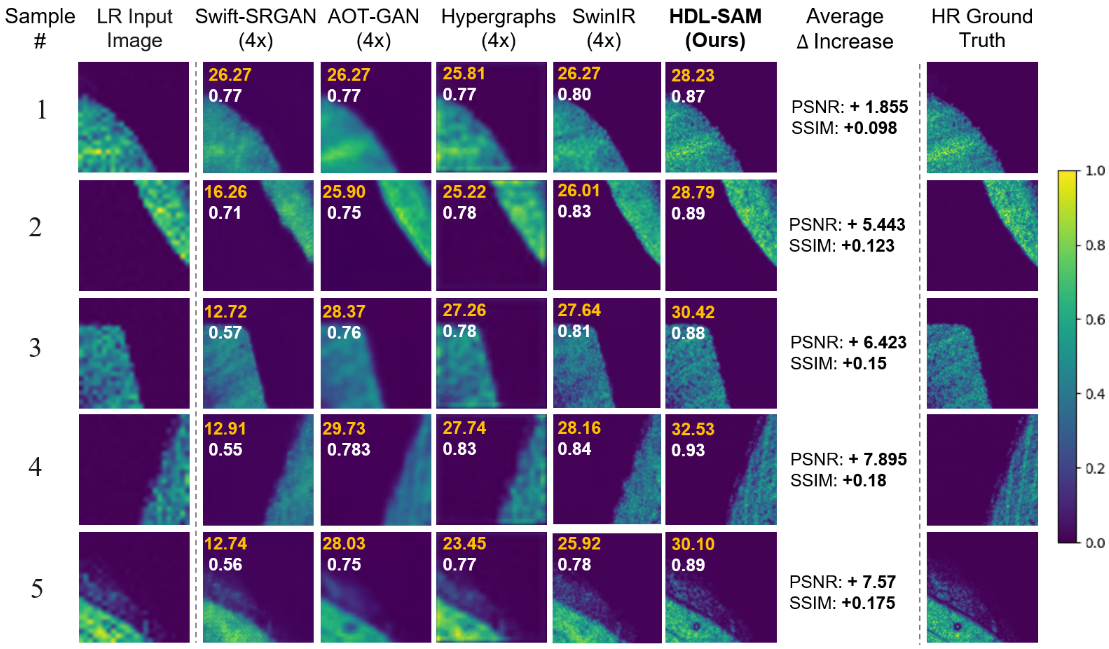

Figure 6. Randomly sampled result (PSNR: orange; SSIM: white) showcases the outputs of benchmarked models trained on acoustic image dataset for visual comparison. The quantitative improvement of HDL-SAM over the combined average performance of other models for each sample is highlighted in the second to last column.

from different image processing tasks to improve performance accuracy on the SAM images. By integrating hypergraphs and SwinIR strategically, HDL-SAM allows us to meet and exceed standard performance criteria such as PSNR and SSIM throughout while ensuring that the improvements are contextually relevant and contribute to the overarching goal of achieving high-quality results with minimal resource investment. Furthermore, it can be seen from the results above that the proposed technique achieves faster convergence on SwinIR instances for the specified task while maintaining high prediction confidence (ref. Figures 4 and 5). This hybrid model capitalizes on the strengths of both architectures, where hypergraphs facilitate detailed image inpainting with a high degree of interpretability, and SwinIR contributes to superior image restoration quality.

More research in this field is important to solve the challenges of handling big image processing tasks and creating strong, reliable solutions that work well and are eco-friendly. By employing transfer learning, we demonstrate sustainable AI practices, reusing pre-trained models to reduce computational resources. This aligns with the goals of decreasing environmental impact and advancing energy-efficient learning. Leveraging hypergraphs and SwinIR, we set a new standard for responsible AI innovation. Further research is critical for overcoming obstacles in large-scale image processing and achieving high-performing, environmentally friendly solutions.

5. Conclusion

In this paper, a hybrid deep learning model named HDL-SAM was introduced, integrating Hypergraphs and SwinIR architectures for image inpainting and restoration tasks, respectively. The model exhibited superior performance, surpassing the current state-of-the-art on the challenging 4× upscaling task. Through extensive experimentation, the efficacy of HDL-SAM was demonstrated by comparing it against individual applications of its component models, as presented in Table 1. The results highlight the significant improvement achieved by this proposed approach. Additionally, Furthermore, HDL-SAM was benchmarked against a range of other models, confirming its superior performance in both quantitative assessments and visual quality. This validation highlights the successful integration of Hypergraphs and SwinIR, leveraging their combined benefits. Overall, this study advances the field of image inpainting and restoration forward, while emphasizing the importance of synergistic integration of diverse deep-learning architectures. HDL-SAM not only sets a new standard for image upscaling tasks, but it also opens up new possibilities for the development of hybrid models in future research.

Acknowledgement

This work was supported by the Research Council of Norway, Cristin Project, ID: 2061348.

Code and Data Availability

The code and data for the paper has been made available at https://github.com/akshitsharma1/HDLSAM

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 6

[2] Pragyan Banerjee, Shivam Milind Akarte, Prakhar Kumar, Muhammad Shamsuzzaman, Ankit Butola, Krishna Agarwal, Dilip K Prasad, Frank Melandsø, and Anowarul Habib. High-resolution imaging in acoustic microscopy using deep learning. *Machine Learning: Science and Technology*, 5(1):015007, 2024. 2, 4

[3] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24, 2009. 2

[4] Andrew Briggs, GAD Briggs, and Oleg Kolosov. *Acoustic microscopy*, volume 67. Oxford University Press, 2010. 2

[5] Jiayin Cai, Changlin Li, Xin Tao, and Yu-Wing Tai. Image multi-inpainting via progressive generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 978–987, 2022. 2

[6] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 4

[7] Haoxuan Che, Siyu Chen, and Hao Chen. Image quality-aware diagnosis via meta-knowledge co-embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19819–19829, 2023. 1

[8] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Cross aggregation transformer for image restoration. *Advances in Neural Information Processing Systems*, 35:25478–25490, 2022. 4

[9] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015. 5

[10] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, 2004. 2

[11] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11065–11074, 2019. 4

[12] Ding Ding, Sundaresh Ram, and Jeffrey J Rodriguez. Perceptually aware image inpainting. *Pattern Recognition*, 83:174–184, 2018. 2

[13] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. 2

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 4

[15] Alexei A Efros and Thomas K Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the seventh IEEE International Conference on Computer Vision*, volume 2, pages 1033–1038, 1999. 1

[16] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3558–3565, 2019. 4

[17] Anowarul Habib and Frank Melands. Chirp coded ultrasonic pulses used for scanning acoustic microscopy. In *2017 IEEE International Ultrasonics Symposium*, pages 1–4, 2017. 2

[18] Anowarul Habib, Amit Shelke, Michael Vogel, Sebastian Brand, Xin Jiang, Ullrich Pietsch, Sourav Banerjee, and Tribikram Kundu. Quantitative ultrasonic characterization of c-axis oriented polycrystalline aln thin film for smart device application. *Acta Acustica United with Acustica*, 101(4):675–683, 2015. 1, 2

[19] Anowarul Habib, Amit Shelke, Michael Vogel, Ullrich Pietsch, Xin Jiang, and Tribikram Kundu. Mechanical characterization of sintered piezo-electric ceramic material using scanning acoustic microscope. *Ultrasonics*, 52(8):989–995, 2012. 1

[20] Anowarul Habib, Juha Vierinen, Ashraful Islam, Inigo Zubiavrre Martinez, and Frank MelandsØ. In vitro volume imaging of articular cartilage using chirp-coded high frequency ultrasound. In *2018 IEEE International Ultrasonics Symposium*, pages 1–4, 2018. 2

[21] Matthias Hofmann, Ralph Pflanzer, Anowarul Habib, Amit Shelke, Jürgen Bereiter-Hahn, August Bernd, Roland Kaufmann, Robert Sader, and Stefan Kippenberger. Scanning acoustic microscopy—a novel noninvasive method to determine tumor interstitial fluid pressure in a xenograft tumor model. *Translational Oncology*, 9(3):179–183, 2016. 2

[22] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018. 5

[23] Yi Jiang, Jiajie Xu, Baoqing Yang, Jing Xu, and Junwu Zhu. Image inpainting based on generative adversarial networks. *IEEE Access*, 8:22884–22892, 2020. 2

[24] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. 2018. 6

[25] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 4

[26] Prakhar Kumar, Nitin Yadav, Muhammad Shamsuzzaman, Krishna Agarwal, Frank Melandsø, and Anowarul Habib.

Numerical method for tilt compensation in scanning acoustic microscopy. *Measurement*, 187:110306, 2022. 6

[27] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023. 4

[28] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 2, 4, 6

[29] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision*, pages 85–100, 2018. 2

[30] Jia Liu, Fang Chen, Xianyu Wang, and Hongen Liao. An edge enhanced srgan for mri super resolution in slice-selection direction. In *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy: 4th International Workshop, MBIA 2019, and 7th International Workshop, MFCA 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings 4*, pages 12–20. Springer, 2019. 4

[31] Ákos Makra, Wolfgang Bost, Imre Kalló, András Horváth, Marc Fournelle, and Miklós Gyöngy. Enhancement of acoustic microscopy lateral resolution: A comparison between deep learning and two deconvolution methods. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(1):136–145, 2019. 4

[32] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*, pages 311–320. Springer, 2018. 4

[33] Ayush Somani, Pragyan Banerjee, Manu Rastogi, Krishna Agarwal, Dilip K Prasad, and Anowarul Habib. Image inpainting with hypergraphs for resolution improvement in scanning acoustic microscopy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3112–3121, 2023. 2

[34] Ayush Somani, Alexander Horsch, and Dilip K Prasad. *Interpretability in deep learning*. Springer, 2023. 2

[35] Yucheng Tang, Riqiang Gao, Ho Hin Lee, Shizhong Han, Yunqiang Chen, Dashan Gao, Vishwesh Nath, Camilo Bermudez, Michael R Savona, Richard G Abramson, et al. High-resolution 3d abdominal segmentation with random patch network fusion. *Medical Image Analysis*, 69:101894, 2021. 2

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[37] Gourav Wadhwa, Abhinav Dhall, Subrahmanyam Murala, and Usman Tariq. Hyperrealistic image inpainting with hypergraphs. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3912–3921, 2021. 2, 4

[38] Mario J Wolf, A Sukumaran Nair, Peter Hoffrogge, Elfgard Kühnicke, and Peter Czurratis. Improved failure analysis in scanning acoustic microscopy via advanced signal processing techniques. *Microelectronics Reliability*, 138:114618, 2022. 2

[39] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. *Advances in Neural Information Processing Systems*, 32, 2019. 4

[40] Siyuan Yan, Zhen Yu, Xuelin Zhang, Dwarikanath Mahapatra, Shekhar S Chandra, Monika Janda, Peter Soyer, and Zongyuan Ge. Towards trustable skin cancer diagnosis via rewriting model's decision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11568–11577, 2023. 1

[41] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018. 2

[42] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4471–4480, 2019. 2

[43] Xianshun Yuan, Xiao Cui, Hui Gu, Mo Wang, Yin Dong, Shifeng Cai, Xiang Feng, and Ximing Wang. Evaluating cervical artery dissections in young adults: a comparison study between high-resolution mri and ct angiography. *The International Journal of Cardiovascular Imaging*, 36(6):1113–1119, 2020. 2

[44] Yuan Zeng and Yi Gong. Nearest neighbor based digital restoration of damaged ancient chinese paintings. In *2018 IEEE 23rd International Conference on Digital Signal Processing*, pages 1–5, 2018. 2

[45] Yuan Zeng, Yi Gong, and Jin Zhang. Feature learning and patch matching for diverse image inpainting. *Pattern Recognition*, 119:108036, 2021. 1

[46] Jiale Zhang, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Accurate image restoration with attention retractable transformer. *arXiv preprint arXiv:2210.01427*, 2022. 4

[47] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In *European conference on computer vision*, pages 649–667. Springer, 2022. 4