Image inpainting in acoustic microscopy

Pragyan Banerjee ⓘ ; Sibasish Mishra ⓘ ; Nitin Yadav ⓘ ; Krishna Agarwal; Frank Melandsø;
Dilip K. Prasad ⓘ ; Anowarul Habib ✉ ⓘ

Check for updates

View Online

Export Citation

CrossMark
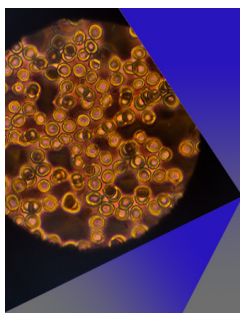
Image inpainting in acoustic microscopy

View Online   Export Citation   CrossMark

Pragyan Banerjee,[1] (ID) Sibasish Mishra,[2] (ID) Nitin Yadav,[2] (ID) Krishna Agarwal,[3] Frank Melandsø,[3]
Dilip K. Prasad,[4] (ID) and Anowarul Habib[3,a] (ID)

AFFILIATIONS

[1] Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati 781039 Assam, India
[2] Department of Physics, Indian Institute of Technology Delhi, Delhi, India
[3] Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway
[4] Department of Computer Science, UiT The Arctic University of Norway, 9037 Tromsø, Norway

[a] Author to whom correspondence should be addressed: anowarul.habib@uit.no

ABSTRACT

Scanning acoustic microscopy (SAM) is a non-ionizing and label-free imaging modality used to visualize the surface and internal structures of industrial objects and biological specimens. The image of the sample under investigation is created using high-frequency acoustic waves. The frequency of the excitation signals, the signal-to-noise ratio, and the pixel size all play a role in acoustic image resolution. We propose a deep learning-enabled image inpainting for acoustic microscopy in this paper. The method is based on training various generative adversarial networks (GANs) to inpaint holes in the original image and generate a 4× image from it. In this approach, five different types of GAN models are used: AOTGAN, DeepFillv2, Edge-Connect, DMFN, and Hypergraphs image inpainting. The trained model's performance is assessed by calculating the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) between network-predicted and ground truth images. The Hypergraphs image inpainting model provided an average SSIM of 0.93 for 2× and up to 0.93 for the final 4×, respectively, and a PSNR of 32.33 for 2× and up to 32.20 for the final 4×. The developed SAM and GAN frameworks can be used in a variety of industrial applications, including bio-imaging.

I. INTRODUCTION

High-frequency Scanning Acoustic Microscopy (SAM) is a highly sensitive and precise technique for imaging the surface and subsurface structures of various materials. It utilizes high-frequency ultrasonic waves to gather information about a specimen, making it a safe way to visualize the interior of objects without physically exposing them. SAM has the capability for non-invasive microstructural characterization of different industrial objects and biological specimens.[1,2] SAM is not only useful for visualizing surface and subsurface structures, but it can also be used to characterize and determine the mechanical properties of piezoelectric materials, structural health monitor (SHM) of composite structures, detect surface defects on polymer circuits, and study the propagation of isotropic or anisotropic phonons.[3–8] The microelectronics and semiconductor industries are highly demanding and competitive markets. In this context, SAM technology plays a critical role in the development of improved mold designs for flip-chip packages.

Additionally, it is capable of handling the complexities involved in miniaturized assemblies, such as chip-scale packages and 3D IC stacks, making it an important tool in the industry.[9,10] The quality of images produced by SAM at a particular frequency depends on the pixel size or scanning steps in both the x and y directions, as well as the spot size of the acoustic beam. Low-resolution images at the same frequency require fewer scanning points, thus reducing the amount of time needed to complete the scanning process. In contrast, high-resolution images at the same frequency require more scanning points, which results in longer data acquisition times. When imaging biological specimens, data acquisition is crucial, and high resolution with smaller step sizes is optimal. However, larger step sizes in scanning can lead to reduced image quality due to less information about the scanned objects. To address this issue, conventional image interpolation or deep learning-based image inpainting techniques can be used to improve image quality.[11–13] To the best of our knowledge, image inpainting in acoustic microscopy is absent from the literature.

17 July 2023 20:27:54

The image inpainting technology is based on deep learning, which has recently emerged as a research hotspot in computer vision.[14–17] Image inpainting is a commonly used technique in computer vision for reconstructing or repairing images. It is used to fix damaged or corrupted areas of an image, remove unwanted elements from photos, and fill in missing parts of an image that are occluded. By utilizing the information already present in an image to infer missing information, this method has addressed the limitations of traditional inpainting methods that have persisted for a long time. As a result, it has significantly improved the quality of the final output. There are two major aspects for generating a contextually plausible and realistic image: (a) global semantic structure and (b) finely detailed texture surrounding the gaps. Image inpainting can be performed by (i) content/texture adapting methods and (ii) generative network-based approaches. The first kind of method uses a simple patching technique of copying the features of surrounding non-hole regions into the holes. Some applications of this kind of patch-matching algorithm that iteratively fills the missing pixels by searching for similar patches among the non-hole pixels in the image are described in Refs. 18–21. However, the limitation of these methods is that they fail to identify the global context or semantics of the image and hence produce implausible results.

There exist various generative network-based approaches for different computer vision applications, such as image super-resolution, image inpainting, image de-blurring, image colorization, etc.[22–30] These approaches use Generative Adversarial Networks (GANs). GANs are an effective approach for generative modeling using deep learning, like convolutional neural networks. It is an unsupervised learning technique that intelligently discovers and learns the features and patterns in inputs and generates outputs that could plausibly belong to the original dataset. The GAN model consists of two main components: a generator that is trained on a dataset to create new examples and a discriminator that distinguishes between the generated examples and the real ones in the dataset. The discriminator helps to ensure that the generator produces plausible outputs that belong to the same domain as the original data. GAN-based approaches have been used for the purpose of image-inpainting for the last several years.[24,31–33] Pathak *et al.* first developed the GAN-based image-inpainting techniques, whose core idea was a channel-wise fully connected layer.[24] This model was considered a baseline for many subsequent models. Iizuka *et al.* uplift the image inpainting to the next level.[34] The model included two types of discriminator networks: a global discriminator network that examined the entire image to ensure overall consistency and a local discriminator network that focused on the details and pixels surrounding the filled hole in the center of the image. This multi-scale discriminator design is adopted by almost all the following image-inpainting papers. The next major breakthrough was in 2018 in the paper "Generative Image Inpainting with Contextual Attention" (also known as DeepFillv1) by Yu *et al.*[14] This model was an improvement over the previous "Shift-Net" model[35] as it used a contextual attention (CA) layer, which is a differential and fully convolutional layer that assigns weights to individual features that show their contribution to each location in the gap/missing region. A subsequent enhanced version of this model was introduced in 2019 in the paper "Free-Form Image Inpainting with Gated Convolution" (also known as DeepFillv2) by Yu *et al.*[25] The most important aspect

of this model is gated convolutions. So far, the focus has been on applying image inpainting methods to optical images, and there is a lack of research on applying these methods to scanning acoustic microscopic images.

The model we utilized in our work is based on the approach presented in "Hyperrealistic Image Inpainting with Hypergraphs" by Wadhwa *et al.* (2021).[36] The model leverages Hypergraph structures to identify and incorporate similar features from the background in order to reconstruct the missing regions. The model consists of two stages, coarse and fine, which aim to produce results that accurately capture the overall context and finer details of the image. This model has the implementation of a trainable method to compute a data-dependent incidence matrix for the Hypergraph convolutions. The local consistency of the image is ensured by gated convolution rather than the regular convolution in the discriminator network.

In this paper, we applied this model to a large dataset of SAM-generated images. Due to variations in image size and aspect ratios, the images were cropped to a uniform size of 96 × 96 pixels. Instead of using an irregular or center mask as described in Ref. 36, we have used a grid mask of alternately white and black pixels for training the model on our dataset. This is performed to "inpaint" every alternate pixel to increase the overall resolution of the image by a factor of 2 times. This process was repeated once more on the modified dataset to achieve a total of four times the resolution. The results of our implementation of the Hypergraph convolution network on the SAM-recorded image dataset are impressive, with structural similarity index measure (SSIM) scores reaching up to 0.93 for 2× and up to 0.93 for the final 4× as well as a peak signal-to-noise ratio (PSNR) of 32.33 for 2× and up to 32.20 for the final 4×, indicating high-quality state-of-the-art image inpainting output. This technique opens up the possibility of removing the barriers imposed by the step-size limitations of high-speed imaging for biological samples. Figure 1 depicts the overall strategy used in this paper.
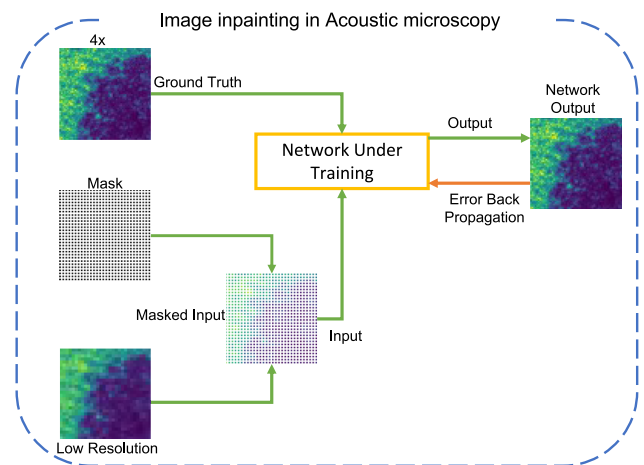

FIG. 1. The overall strategy used in this paper. We use the alternate hole mask to create an input image for the model and create a high-resolution image with the help of image inpainting.

II. METHOD

Image inpainting refers to the task of reconstructing missing regions of an image. This work aims to use image inpainting techniques for image high-resolution. This has been performed by employing mask creation. The mask used for training the dataset is a matrix of black pixels for every three white pixels. The lower resolution image is increased in size by four times by incorporating three white pixels in between each recorded data point. Training and using the model directly for 4× inpainting does not give state-of-the-art results. Hence, the model was first trained to perform a 2× up-sampling followed by a subsequent 2× up-sampling to provide an overall 4× up-sampled result. For this, the approach to mask creation was similar. Initially, the lower resolution image is increased two times by incorporating one white pixel between each recorded data point. Then the white pixels are filled using image inpainting techniques. This process is repeated to achieve the resultant 4× image. For this purpose, various image inpainting approaches, namely, AOTGAN,[21] DeepFillv2,[37] Edge-Connect,[26] DMFN[38] and Hypergraphs image inpainting,[39] have been implemented. These were initially tested using the CelebA-HQ dataset[40] and pre-trained models. It was observed that the Hypergraphs image inpainting technique provided the best results for our approach. Hence, Subsection II A describes the Hypergraph image inpainting architecture.

A. Hypergraphs

A Hypergraph, defined by $G = (V, E, W)$, consists of hyperedges that connect two or more vertices. Here, $V = v_1, \ldots, v_n$ is the set of vertices, $E = e_1, \ldots, e_n$ is the set of hyperedges, and $W \in \mathbb{R}^{M \times M}$ is a diagonal matrix containing the weight of each matrix. The Hypergraph G can also be defined by the incidence matrix $H \in \mathbb{R}^{N \times M}$, which is defined as follows:

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

Given a Hypergraph G, vertex degree D, $D \in \mathbb{R}^{N \times N}$, and hyperedge degree $B \in \mathbb{R}^{M \times M}$, we have

$$D_{ii} = \sum_{e=1}^{M} W_{ee} H_{ie}, \tag{1}$$

$$B_{ee} = \sum_{i=1}^{N} H_{ie}. \tag{2}$$

In addition, the normalized Hypergraph Laplacian matrix, $\Delta \in \mathbb{R}^{N \times N}$, is given as

$$\Delta = I - D^{\frac{1}{2}} H B^{-1} H^T D^{-\frac{1}{2}}. \tag{3}$$

The above matrix is symmetric positive semi-definite,[41] and its Eigen decomposition is given by

$$\Delta = \Phi \Lambda \Phi^T. \tag{4}$$

The above decomposition is used to get the orthogonal eigenvectors $\Phi = \{\phi_1, \ldots, \phi_N\}$ and diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$, which contains the corresponding non-negative eigenvalues. The

Hypergraph Fourier transform is given by $\hat{x} = \Phi^T x$. The convolution on the signal $x \in \mathbb{R}^N$ is defined as

$$g \odot x = \Phi g(\Lambda) \Phi_T x, \tag{5}$$

where $g(\Lambda) = \text{diag}(g(\lambda_1), \ldots, g(\lambda_N))$ is a function of Fourier coefficients. Parameterizing $g(\Lambda)$ with truncated Chebyshev polynomials up to Kth order, the convolution operation on the Hypergraph signal can be defined as

$$g \odot x = \sum_{k=0}^{K} \theta_k T_k(\Delta) x, \tag{6}$$

$$g \odot x = \theta D^{-\frac{1}{2}} H W B^{-1} H^T D^{-\frac{1}{2}} x. \tag{7}$$

For a given Hypergraph signal $X^l \in \mathbb{R}^{N \times C_t}$, where C_t is the dimension of the feature vector l, the convolution operation can be generalized to the multi-layer Hypergraph convolution network as

$$X^{l+1} = \sigma\left(D^{-\frac{1}{2}} H W B^{-1} H^T D^{-\frac{1}{2}} X^l \Theta\right), \tag{8}$$

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

B. Hypergraphs convolution on spatial features

Simple graphs can be considered a special case of Hypergraphs, where each hyperedge connects only two nodes. They can easily represent the pair-wise relationship among data, but it is difficult to represent the spatial features and their relationship in an image, which is why Hypergraphs are used instead of graphs. To transform the spatial features $F^l \in \mathbb{R}^{hw \times c}$ into the graph-like structure, each spatial feature is considered as a node having a feature vector of dimension c, $X^l \in \mathbb{R}^{hw \times c}$.

For the incidence matrix H, instead of using the Euclidean distance[42,43] between features of images, cross-correlation of the spatial features is used to calculate each node's contribution to the hyperedge. Therefore, we have

$$H = \Psi(X) \Lambda(X) \Psi(X)^T \Omega(X), \tag{9}$$

where $\Psi(X) \in \mathbb{R}^{N \times C}$ is the linear embedding of the input features followed by the ReLU activation function, and $\hat{C}$ is the dimension of the feature vector after the linear embedding. $\lambda(X) \in \mathbb{R}^{\hat{C} \times \hat{C}}$ is a diagonal matrix that helps in learning a better distance metric among the nodes for the incidence matrix H, $\Omega(X) \in \mathbb{R}^{N \times M}$ helps to determine the contribution of each node for each hyperedge, and m is the number of hyperedges in the Hypergraph. $\Psi(X)$ is implemented by 1×1 convolution on the input features; $\Lambda(X)$ is implemented by channel-wise global average pooling followed by a 1×1 convolution, as in Ref. 44; and $\Omega(X)$ is implemented using a 7×7 filter.

Hence, we get

$$H^l = \Psi(X^l) \Lambda(X^l) \Psi(X^l)^T \Lambda(X^l)^T, \tag{10}$$

$$\Psi(X^l) = conv(X^l, W_\Psi^l), \tag{11}$$

$$\Lambda(X^l) = \text{diag}\left(conv(\hat{X}^l, W_\Lambda^l)\right), \tag{12}$$

$$\Omega(X^l) = conv(X^l, W_\Omega^l), \tag{13}$$

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

$$X^{l+1} = \sigma(\Delta X^l \Theta), \tag{14}$$

where $\Theta \in \mathbb{R}^{C_l \times c_{l+1}}$ is the learnable parameter, and σ is the ELU[45] non-linear activation function.

C. Architecture and training parameters

Figure 2 depicts the architecture of the Hypergraphs image inpainting model. The Hypergraphs image inpainting network consists of a two-stage coarse-to-fine network architecture. While the coarse network roughly fills in the missing region, which gets naively blended with the input image, the refine network predicts the finer results with sharp edges. Hypergraph layers with high-level feature maps are used in the refine layer to increase the receptive field of our network and obtain distant global information about the image. Dilated convolutions[34] are used to further expand the receptive fields of the coarse and refine networks. In addition, gated convolutions[37] are used to improve performance, which can be defined as

$$Gating = Conv(W_g, I), \tag{15}$$

$$Features = Conv(W_f, I), \tag{16}$$

$$O = \phi(Features) \odot \sigma(Gating), \tag{17}$$

where W_g and W_f are two different learnable parameters for convolution operations, σ is the sigmoid activation function, and ϕ is a non-linear activation function, such as ReLU, ELU, and LeakyReLU. In addition, to prevent deterioration of the color coherency of the completed image, batch normalization is removed.[34] In this method, the discriminator has an architecture similar to that of the Patch-GAN.[46] All batch normalization layers are removed, and all convolution layers are replaced with a gated convolution, which enforces local consistency in the completed image. The discriminator is provided with both a mask and a completed/original image.

For an input image I_{in} with holes and a binary mask R (with 1 for holes), the network predicts I_{coarse} and I_{refine} from the coarse and refine networks, respectively. For the corresponding ground truth I_{gt}, the model is trained on a combination of content loss, adversarial loss, perpetual loss, and edge loss. $L1$ loss is used on both I_{coarse} and
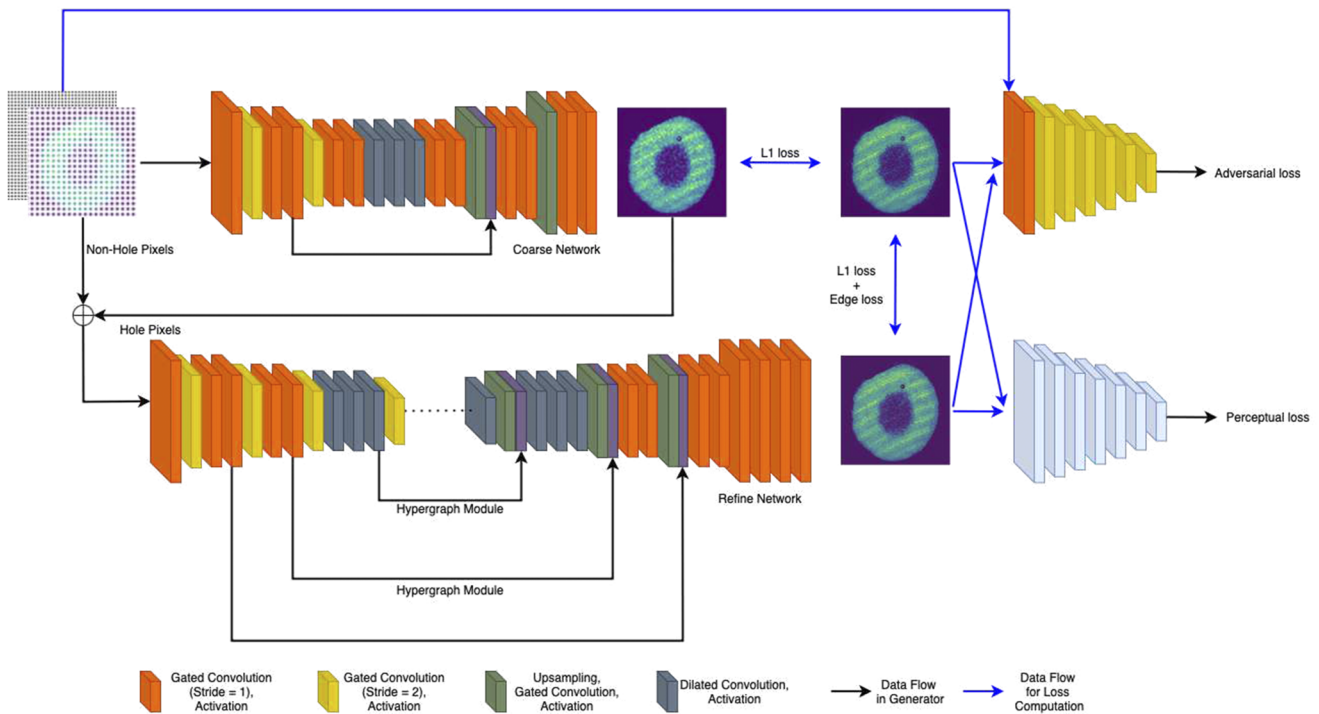

FIG. 2. The architecture of the Hypergraphs image inpainting model. Its network consists of a two-stage coarse-to-fine network architecture. The coarse network roughly fills the missing region, which gets naively blended with the input image, and the refined network predicts the finer results with sharp edges.

I_{refine} to maintain pixel level consistency. Hence, the content loss is defined as

$$L_{hole} = \|R \odot (I_{refine} - I_{gt})\|_1 + \frac{1}{2}\|R \odot (I_{coarse} - I_{gt})\|_1, \quad (18)$$

$$L_{valid} = \|(1-R) \odot (I_{refine} - I_{gt})\|_1 + \frac{1}{2}\|(1-R) \odot (I_{coarse} - I_{gt})\|_1, \quad (19)$$

where L_{hole} is the loss for the hole pixel values, and L_{valid} is the loss for the non-pixel values. The adversarial loss, which has been shown to generate realistic and globally consistent images, can be formulated as a min–max problem,

$$L_{GAN} = \max_D \min_G \mathbb{E}[\log(D(I_{gt}, R))] + \mathbb{E}[\log(1 - D(G(I_{in}), R)], \quad (20)$$

where G denotes the image inpainting network, which predicts the final image I_{refine}, and D is the discriminator. For a given input x, let $\phi_l(x)$ denote the high-dimension features of the lth activation layer of the pre-trained network, and then the perpetual loss is defined as

$$L_p = \sum_l \|\phi_l(G(I_{im})) - \phi_l(I_{gt})\|_1. \quad (21)$$

The perceptual loss for the final prediction I_{refine} and I_{comp} are computed, where I_{comp} is the final prediction, but the non-hole pixels are set directly to ground truth.[46] Edge-preserving loss[47] is used to maintain edges in the predicted images, which can be defined as

$$L_{edge} = \|E(I_{refine}) - E(I_{gt})\|_1, \quad (22)$$

where $E(\cdot)$ is the Sobel filter. Hence, the total loss L_{total} can be written as

$$L_{total} = \lambda_{hole}L_{hole} + \lambda_{valid}L_{valid} + \lambda_{adv}L_{adv} + \lambda_p L_p + \lambda_{edge}L_{edge}, \quad (23)$$

where λ_{hole}, λ_{valid}, λ_{adv}, λ_p, and λ_{edge} are the weights for hole, valid, adversarial, perceptual, and edge loss, respectively.
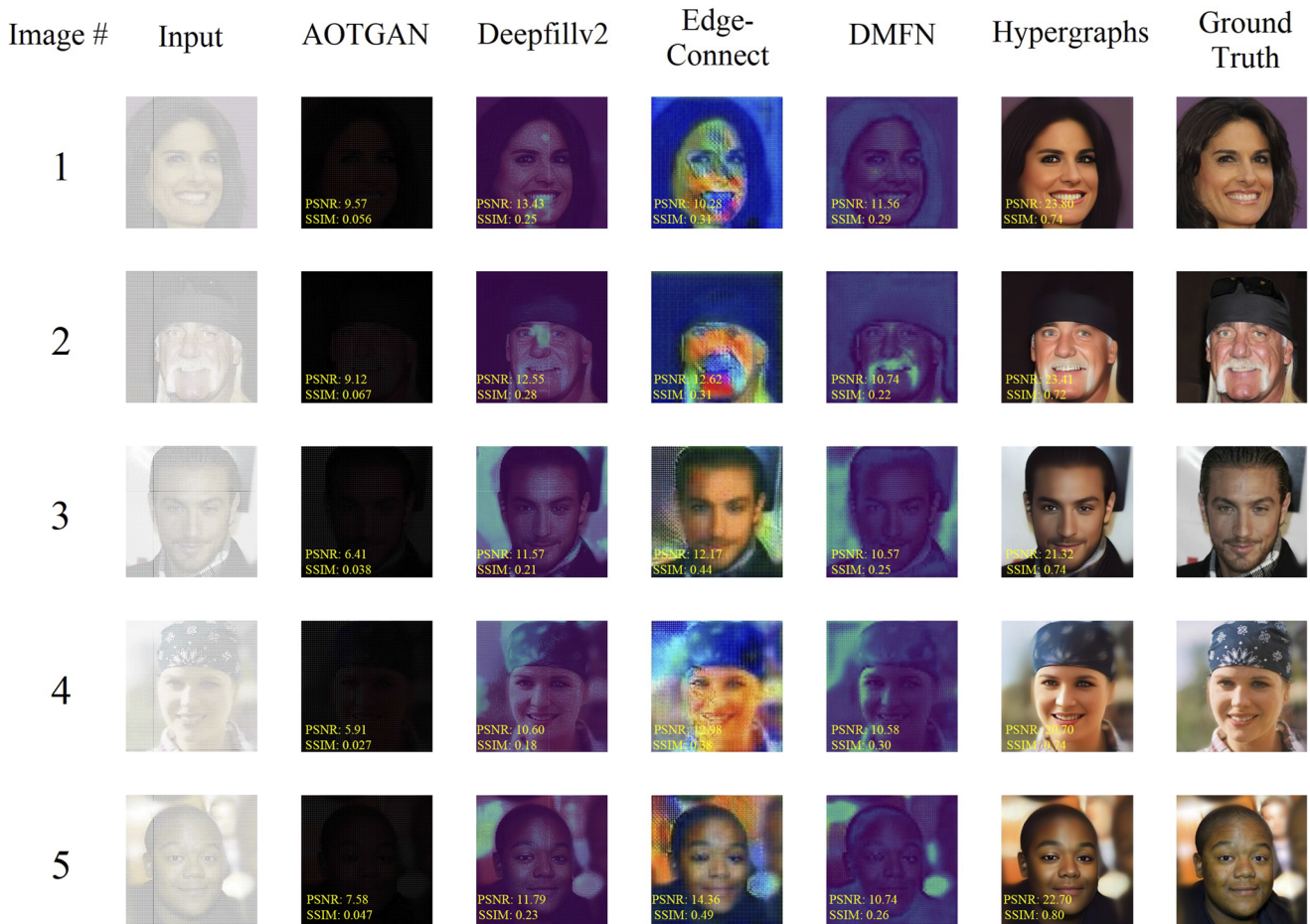

FIG. 3. The outputs of some of the CelebA-HQ dataset for all the models. (Portions of this page are modifications based on work created and shared by Google and used according to terms described in the Creative Commons 4.0 Attribution License.)[48,49]

TABLE I. The SSIM and PSNR scores of various models tested our dataset using the method stated in this paper.

Model	AOT-GAN	DeepFillv2	Edge-connect	DMFN	Hypergraphs
PSNR	7.72	11.99	12.48	10.84	22.36
SSIM	0.047	0.23	0.39	0.26	0.70

TABLE II. The SSIM scores of various models tested on the CelebA-HQ dataset using the method stated in this paper.

Image no.	AOT-GAN	DeepFillv2	Edge-connect	DMFN	Hypergraphs
1	0.056	0.25	0.31	0.29	0.74
2	0.067	0.28	0.31	0.22	0.72
3	0.038	0.21	0.44	0.25	0.74
4	0.027	0.18	0.38	0.30	0.74
5	0.047	0.23	0.49	0.26	0.80

TABLE III. The PSNR scores of various models tested on the CelebA-HQ dataset using the method stated in this paper.

Image no.	AOT-GAN	DeepFillv2	Edge-connect	DMFN	Hypergraphs
1	9.57	13.43	10.28	11.56	23.80
2	9.12	12.55	12.62	10.74	23.41
3	6.41	11.57	12.17	10.57	21.32
4	5.91	10.60	12.98	10.58	20.70
5	7.58	11.79	14.36	10.74	22.70

III. RESULTS ON THE CELEBA-HQ DATA SET

Initially, all the models were trained on the CelebA-HQ dataset as a proof of concept using the masking and training strategy defined in 2. Figure 3 shows the outputs of some of the CelebA-HQ dataset for all the models.[48,49] A PSNR score of 22.94 and an SSIM score of 0.79 were obtained. The mean SSIM and PSNR scores of all the models in the CelebA-HQ dataset are shown in Table I. The following are some examples of the results obtained, as shown in Fig. 5 and Tables II and III.

IV. EXPERIMENTAL DATASET

We used a high-resolution technique in scanning acoustic microscopy to create 4× high-resolution images from low-resolution images. There are a total of 33 high-resolution images recorded using scanning acoustic microscopy. These images are measured with a step size of 50 μm. The images are of various sizes and aspect ratios. Each image is cropped into multiple images to create diversity and, hence, more robust training of the network. Cropping is performed by starting from the image's top-left corner and striding in the G and H directions. The resulting cropped image has a dimension of 96 × 96 pixels. This was performed to maintain a uniform size for training and ensure that the images' overall semantics are somewhat preserved. The training dataset consisted of 402 such 96 × 96 images. During the training, the image's values were scaled in the range [0–1]. The motivation for choosing a mask is inspired by the fact that the scanning acoustic microscope is operated at two different step

sizes. The low-resolution images are recorded as having a step size of 200 μm in contrast to the 50 μm step size of the higher-resolution images used for training the model.

V. MATERIALS AND METHODS

A. Experimental setup

The scanning acoustic microscope (SAM) has two operational modes: reflection and transmission. Figure 4 presents a labeled image of SAM that is utilized for image acquisition. Furthermore, details regarding the working principles of these modes can be found elsewhere. In this article, we have focused on the reflection mode
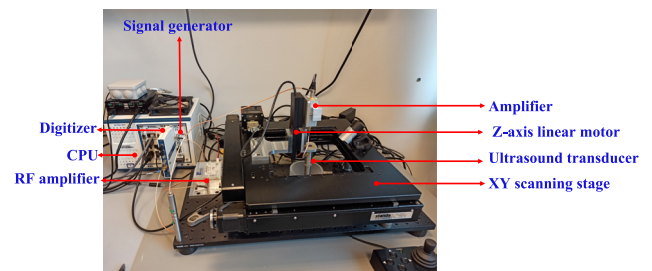

FIG. 4. A labeled image of SAM that is utilized for image acquisition. The experimental setup demonstrates all the fundamental components that constitute a SAM.
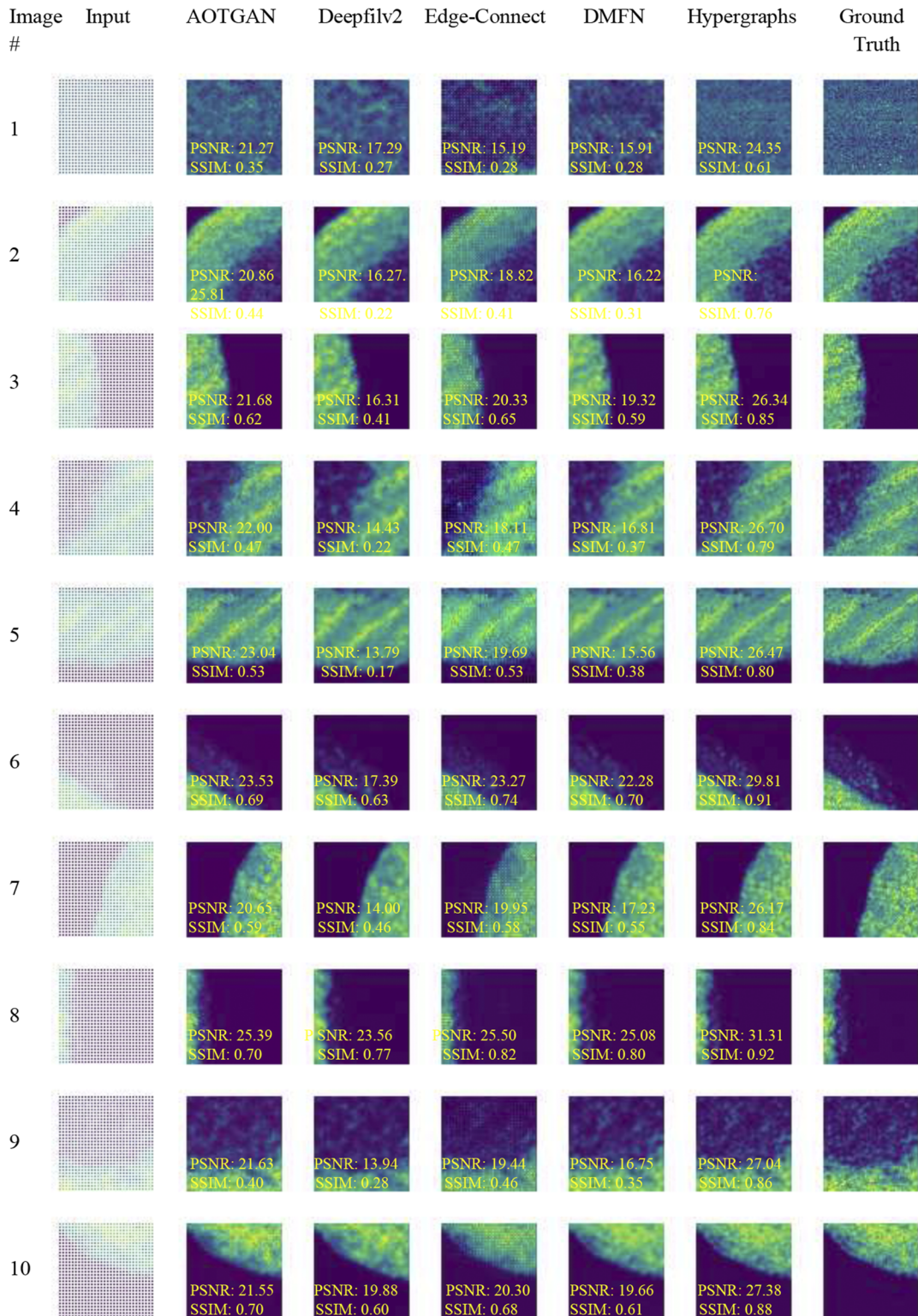
FIG. 5. The output images of ten input images from the scanning acoustic microscopy dataset and the corresponding outputs of the five models using our super-resolution method.

to scan the samples. To focus acoustic energy through a coupling medium (in this case, water), a concave spherical sapphire lens rod is commonly used. Next, ultrasound signals are generated from a signal generator and transmitted to the sample. The reflected waves are then recorded, and the resulting digitized signal from the sample is referred to as an A-scan or amplitude scan. To obtain a C-scan of the sample, this process is repeated at various points in the XY plane. Alternatively, a C-scan can be viewed as the summation of A-scans in two dimensions.

The experimental data were collected using a custom-built SAM (shown in Fig. 5) that included a high-precision scanning stage from Standa (8MTF-200, Motorized XY Microscope Stage) and was controlled by a LabVIEW program.[50] Previous work from the same group utilized a similar experimental setup to correct for inclined samples.[51] The acoustic imaging features were implemented using National Instruments' PXIe FPGA modules and FlexRIO hardware, which were housed in a PXIe chassis (PXIe-1082) that included an arbitrary waveform generator (AT-1212).[52] The transducer was excited with Mexican hat signals and transmitted through an RF amplifier (AMP018032-T) to amplify the ultrasonic signals. The resulting acoustical reflections from the sample surface were then amplified with a custom-designed amplifier, and these signals were further amplified using a custom-designed pre-amplifier and digitized with a 12-bit high-speed (1.6 GS/s) digitizer (NI-5772). For ground truth, a 50 MHz focused transducer manufactured by Olympus was employed, featuring a 6.35 mm aperture and a 12 mm focal length. The transducer was used to scan both the coin and the biological specimen. During scanning, the acoustic energy was focused on the top surface of the coin, and the sample was scanned in the x and y directions with 50 μm steps. Low-resolution images were obtained using a 20 MHz transducer with a focal length of 50 mm. All experiments were conducted in distilled water while maintaining a constant room temperature of ~22 °C. To evaluate the models, a discarded reindeer antler collected from the jungle in Tromsø, Norway, was used as a biological sample for imaging. Prior to scanning, the moss on the antler was removed by cleaning it with

TABLE IV. The SSIM and PSNR scores of various models tested our dataset using the method stated in this paper.

Model	AOT-GAN	DeepFillv2	Edge-connect	DMFN	Hypergraphs
PSNR	22.47	18.09	20.64	19.46	27.96
SSIM	0.56	0.43	0.54	0.50	0.82

TABLE V. The SSIM scores of various models tested on our dataset using the method stated in this paper.

Image no.	AOT-GAN	DeepFillv2	Edge-connect	DMFN	Hypergraphs
1	0.35	0.27	0.28	0.27	0.61
2	0.44	0.22	0.41	0.31	0.76
3	0.62	0.41	0.65	0.59	0.85
4	0.47	0.22	0.47	0.37	0.79
5	0.53	0.17	0.53	0.38	0.80
6	0.69	0.63	0.74	0.70	0.91
7	0.59	0.46	0.58	0.55	0.84
8	0.70	0.77	0.82	0.80	0.92
9	0.40	0.28	0.46	0.35	0.86
10	0.70	0.60	0.68	0.61	0.88

TABLE VI. The PSNR scores of various models tested on our dataset using the method stated in this paper.

Image no.	AOT-GAN	DeepFillv2	Edge-connect	DMFN	Hypergraphs
1	21.27	17.29	15.19	15.91	24.35
2	20.86	16.27	18.82	16.22	25.81
3	21.68	16.31	20.33	19.32	26.34
4	22.00	14.43	18.11	16.81	26.70
5	23.04	13.79	19.69	15.56	26.47
6	23.53	17.39	23.27	22.28	29.81
7	20.65	14.00	19.95	17.23	26.17
8	25.39	23.56	25.50	25.08	31.31
9	21.63	13.94	19.44	16.75	27.04
10	21.55	19.88	20.30	19.66	27.38

lukewarm water and 96% ethanol. Next, the sample was diced and boiled in distilled water at 100 °C to eliminate any undesired biological substances from the antler. Finally, the sample was placed on the sample holder and allowed to dry before being scanned.

VI. RESULTS AND DISCUSSION

In this work, the model was successfully able to upsample a low-resolution (bigger step size) image four times using our mask. The SSIM and PSNR scores of all five models in our dataset are given in Table IV.

As stated earlier, out of the five models tried, the Hypergraphs image inpainting model gave the best results, as evident from the SSIM and PSNR scores given in Tables V and VI as well as a visual inspection in Fig. 6. A PSNR score of 27.96 and an SSIM score of 0.8234 were obtained using the Hypergraphs model. In our comparison, AOT-GAN was second best, but both the SSIM and PSNR scores for AOT-GAN are very low as compared to the Hypergraphs
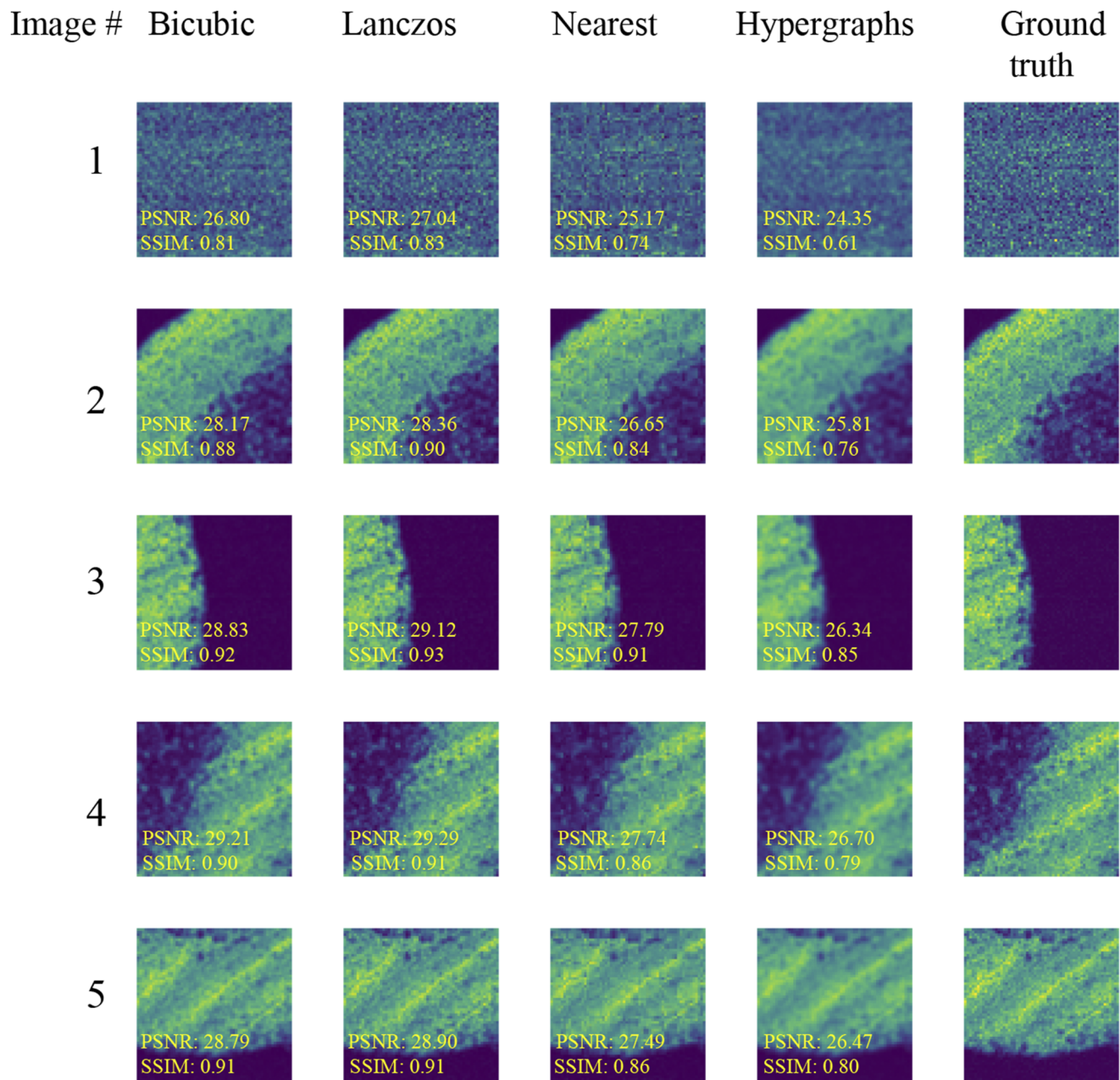

FIG. 6. Some of the outputs of the Hypergraphs image and the corresponding output from conventional digital resolution enhancement techniques.

model. Hence, the Hypergraphs model was best able to learn to fill up the missing pixels in our mask and generate 4× larger images.

Although by visual inspection, the resultant images are better than the CelebA-HQ dataset results. The resultant images are more comparable between the models in this case than for the CelebA-HQ dataset. This might be because all the images in our dataset have many similar characteristics. However, a closer look will reveal that indeed Hypergraphs model gave results that were the closest to the ground truth as compared to other models that we tried.

VII. COMPARISON WITH CONVENTIONAL DIGITAL RESOLUTION ENHANCEMENT TECHNIQUES

We compared the results obtained in this paper with some conventional digital resolution enhancement techniques: bicubic, Lanczos, and nearest neighbor. The SSIM and PSNR scores have been shown in Table VII.

Some of the results are shown in Fig. 6, and their SSIM and PSNR scores have been tabulated in Tables VIII and IX, respectively. It can be seen from the results that the results of the Hypergraphs model were very close to those of the conventional techniques. However, it should be noted that the main purpose of the work was to compare AI-based inpainting models with our new grid mask, which can mimic super-resolution to a large extent.

TABLE VII. The SSIM and PSNR scores of various models tested on the CelebA-HQ dataset using the method stated in this paper.

Method	Bicubic	Lanczos	Nearest neighbor	Hypergraphs
PSNR	30.34	30.54	28.87	27.96
SSIM	0.91	0.92	0.88	0.82

TABLE VIII. The SSIM scores of various models tested on the CelebA-HQ dataset using the method stated in this paper.

Image no.	Bicubic	Lanczos	Nearest neighbor	Hypergraphs
1	0.81	0.83	0.74	0.61
2	0.88	0.90	0.84	0.76
3	0.92	0.93	0.91	0.85
4	0.90	0.91	0.86	0.79
5	0.91	0.91	0.86	0.80

TABLE IX. The PSNR scores of various models tested on the CelebA-HQ dataset using the method stated in this paper.

Image no.	Bicubic	Lanczos	Nearest neighbor	Hypergraphs
1	26.80	27.04	25.17	24.35
2	28.17	28.36	26.65	25.81
3	28.83	29.12	27.79	26.34
4	29.21	29.29	27.74	26.70
5	28.79	28.90	27.49	26.47

VIII. CONCLUSION

In this paper, we have developed an acoustic microscopy system assisted with deep learning to improve the image resolution of industrial and biological samples with the help of image inpainting. Deep learning was employed to fill the gaps in the SAM images. This work compared five deep learning models: AOT GAN, DeepFillv2, Edge-Connect, DMFN, and Hypergraphs. All five models were trained and tested on test images, and the results in terms of PSNR and SSIM scores were compared. The Hypergraphs model presented the best results in terms of SSIM and PSNR. The Hypergraphs image inpainting network consists of a two-stage coarse-to-fine network architecture. While the coarse network roughly fills in the missing region, which gets naively blended with the input image, the refine network predicts the finer results with sharp edges. Transfer learning was considered in the process. This was to prevent the model from over-fitting, as limited (800) images were used for training purposes.

In this work, we made a mask of alternate data points and white pixels. Later on, we filled those white points to create a 2× image. By repeating this process again on the 2× image, we were able to create a 4× image. In this process, the Hypergraphs model was the most successful, giving SSIM and PSNR scores of 0.70 and 22.36, respectively, on the CelebA-HQ dataset. Finally, on our acoustic scan dataset, we were able to achieve a mean SSIM score of 27.96 and a mean PSNR score of 0.82. Hence, it is found that even with a minimal training dataset, deep learning-based models, specifically the Hypergraphs model, can provide satisfactory results.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

A.H., D.K.P., and K.A. have conceptualized the idea. A.H. designed the experiments. P.B. implemented the deep learning approach with initial help from S.M. F.M. developed the SAM. N.Y. performed the SAM experiments with the help of A.H. Funding was secured by F.M. and A.H. Formal analysis and experimental validation were performed by P.B., who also wrote the original draft, reviewed, and edited the article with support from all co-authors.

Pragyan Banerjee: Formal analysis (equal); Methodology (equal); Validation (equal); Writing – original draft (equal). **Sibasish Mishra**: Formal analysis (equal); Methodology (equal); Writing –

review & editing (equal). **Nitin Yadav**: Data curation (equal); Writing – review & editing (equal). **Krishna Agarwal**: Conceptualization (equal); Investigation (equal); Project administration (equal); Writing – review & editing (equal). **Frank Melandsø**: Methodology (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Dilip K. Prasad**: Conceptualization (equal); Formal analysis (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Anowarul Habib**: Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. Briggs, G. Briggs, and O. Kolosov, *Acoustic Microscopy* (Oxford University Press, 2010), Vol. 67.

[2] M. Hofmann, R. Pflanzer, A. Habib, A. Shelke, J. Bereiter-Hahn, A. Bernd, R. Kaufmann, R. Sader, and S. Kippenberger, "Scanning acoustic microscopy—A novel noninvasive method to determine tumor interstitial fluid pressure in a xenograft tumor model," Transl. Oncol. **9**, 179–183 (2016).

[3] A. Habib, A. Shelke, M. Vogel, S. Brand, X. Jiang, U. Pietsch, S. Banerjee, and T. Kundu, "Quantitative ultrasonic characterization of c-axis oriented polycrystalline AlN thin film for smart device application," Acta Acust. Acust. **101**, 675–683 (2015).

[4] A. Habib and F. Melands, "Chirp coded ultrasonic pulses used for scanning acoustic microscopy," in *2017 IEEE International Ultrasonics Symposium (IUS)* (IEEE, 2017), pp. 1–4.

[5] A. Habib, A. Shelke, M. Vogel, U. Pietsch, X. Jiang, and T. Kundu, "Mechanical characterization of sintered piezo-electric ceramic material using scanning acoustic microscope," Ultrasonics **52**, 989–995 (2012).

[6] S. Wagle, A. Habib, and F. Melandsø, "Ultrasonic measurements of surface defects on flexible circuits using high-frequency focused polymer transducers," Jpn. J. Appl. Phys. **56**, 07JC05 (2017).

[7] B. Tittmann, C. Miyasaka, M. Guers, H. Kasano, and H. Morita, "Non-destructive evaluation (NDE) of aerospace composites: Acoustic microscopy," in *Non-Destructive Evaluation (NDE) of Polymer Matrix Composites* (Elsevier, 2013), pp. 423–449e.

[8] A. Habib, J. Vierinen, A. Islam, I. Z. Martinez, and F. MelandsØ, "*In vitro* volume imaging of articular cartilage using chirp-coded high frequency ultrasound," in *2018 IEEE International Ultrasonics Symposium (IUS)* (IEEE, 2018), pp. 1–4.

[9] M. Wolf, A. S. Nair, P. Hoffrogge, E. Kühnicke, and P. Czurratis, "Improved failure analysis in scanning acoustic microscopy via advanced signal processing techniques," Microelectron. Reliab. **138**, 114618 (2022).

[10] M. Wolf, P. Hoffrogge, E. Kühnicke, P. Czurratis, and C. Kupsch, "Inspection of multilayered electronic devices via scanning acoustic microscopy using synthetic aperture focusing technique," in *2022 IEEE International Ultrasonics Symposium (IUS)* (IEEE, 2022), pp. 1–4.

[11] M. Nitzberg, D. Mumford, and T. Shiota, *Filtering, Segmentation and Depth* (Springer, 1993), Vol. 662.

[12] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH, 2000), pp. 417–424.

[13] Z. Qin, Q. Zeng, Y. Zong, and F. Xu, "Image inpainting based on deep learning: A review," Displays **69**, 102028 (2021).

[14] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018).

[15] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," International Conference on Learning Representations (2021).

[16] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European conference on computer vision (ECCV)* (ECCV, 2018), pp. 85–100.

[17] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2018), pp. 9446–9454.

[18] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," IEEE Trans. Image Process. **13**, 1200–1212 (2004).

[19] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," ACM Trans. Graphics **22**, 303–312 (2003).

[20] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," ACM Trans. Graphics **28**, 24 (2009).

[21] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the Seventh IEEE International Conference on Computer Vision* (IEEE, 1999), Vol. 2, pp. 1033–1038.

[22] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2017).

[23] K. Nazeri, H. Thasarathan, and M. Ebrahimi, "Edge-informed single image super-resolution," IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (2019).

[24] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).

[25] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2019).

[26] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "Edgeconnect: Generative image inpainting with adversarial edge learning," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops (2019).

[27] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2018).

[28] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," IEEE International Conference on Computer Vision (ICCV) (2016).

[29] W. Quan, R. Zhang, Y. Zhang, Z. Li, J. Wang, and D.-M. Yan, "Image inpainting with local and global refinement," IEEE Trans. Image Process. **31**, 2405–2420 (2022).

[30] H. Xiang, Q. Zou, M. A. Nawaz, X. Huang, F. Zhang, and H. Yu, "Deep learning for image inpainting: A survey," Pattern Recognit. **134**, 109046 (2023).

[31] Y. Jiang, J. Xu, B. Yang, J. Xu, and J. Zhu, "Image inpainting based on generative adversarial networks," IEEE Access **8**, 22884–22892 (2020).

[32] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," arXiv:1803.07422 (2018).

[33] J. Cai, C. Li, X. Tao, and Y.-W. Tai, "Image multi-inpainting via progressive generative adversarial networks," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, 2022), pp. 978–987.

[34] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," ACM Trans. Graphics **36**, 107 (2017).

[35] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-net: Image inpainting via deep feature rearrangement," Computer Vision – ECCV 2018 **11218** (2018).

[36] G. Wadhwa, A. Dhall, S. Murala, and U. Tariq, "Hyperrealistic image inpainting with hypergraphs," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2021), pp. 3911–3920.

[37]Y. Zeng, J. Fu, H. Chao, and B. Guo, "Aggregated contextual transformations for high-resolution image inpainting," IEEE Transactions on Visualization and Computer Graphics (2022).

[38]Z. Hui, J. Li, X. Wang, and X. Gao, "Image fine-grained inpainting," arXiv:2002.02609 (2020).

[39]G. Wadhwa, A. Dhall, S. Murala, and U. Tariq, "Hyperrealistic image inpainting with hypergraphs," in 2021 IEEE Winter Conference on Applications of Computer Vision (WACV) (IEEE, 2021), pp. 3912–3921.

[40]T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," ICLR 2018 (2018).

[41]B. Schölkopf, J. Platt, and T. Hofmann, *Learning with Hypergraphs: Clustering, Classification, and Embedding* (NIPS, 2007), pp. 1601–1608.

[42]Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," Proceedings of the AAAI Conference on Artificial Intelligence (2019).

[43]N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar, "HyperGCN: A new method of training graph convolutional networks on hypergraphs," Proceedings of the 33rd International Conference on Neural Information Processing Systems (2019).

[44]J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018).

[45]D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 4th International Conference on Learning Representations, {ICLR} 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings (2016).

[46]P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2017), pp. 5967–5976.

[47]R. K. Pandey, N. Saha, S. Karmakar, and A. G. Ramakrishnan, "MSCE: An edge preserving robust loss function for improving super-resolution algorithms," Springer, Cham (2018).

[48]Tensorflow, High-quality version of the CelebA dataset, consisting of 30 000 images in 1024 × 1024 resolution, https://www.tensorflow.org/datasets/catalog/celeb_a_hq; accessed 14 March 2023.

[49]tkarras (2023). "Progressive growing of GANs for improved quality, stability, and variation." Github. https://github.com/tkarras/progressive_growing_of_gans; accessed 14 March 2023.

[50]Standa, Motorized xy microscope stage-motorized positioners and controllers-catalog-opto-mechanical products-standa, https://www.standa.lt/products/catalog/motorised_positioners?item=609&prod=motorized_xy_microscope_stage, 2022, accessed 16 March 2022.

[51]P. Kumar, N. Yadav, M. Shamsuzzaman, K. Agarwal, F. Melandsø, and A. Habib, "Numerical method for tilt compensation in scanning acoustic microscopy," Measurement **187**, 110306 (2022).

[52]Becker, Becker nachrichtentechnik gmbh (bnt), "1 W wideband Amplifier," https://www.becker-rf.com/files_db/1608213582_2017_17.pdf, 2022, accessed 16 March 2022.

17 July 2023 20:27:54