Image Inpainting with Hypergraphs for Resolution Improvement in Scanning Acoustic Microscopy

Ayush Somani[1,*]
UiT The Arctic University of Norway
[1]{firstname.lastname}@uit.no

Pragyan Banerjee[2,*]
Indian Institute of Technology Guwahati, India
[2]{firstname.lastname}@iitg.in

Manu Rastogi[3]
Advanced Micro Devices, Inc.
[3]{firstname.lastname}@amd.com

Anowarul Habib[1], Krishna Agarwal[1], Dilip K. Prasad[1]
UiT The Arctic University of Norway
[1]{firstname.lastname}@uit.no

Abstract

Scanning Acoustic Microscopy (SAM) uses high-frequency acoustic waves to generate non-ionizing, label-free images of the surface and internal structures of industrial objects and biological specimens. The resolution of SAM images is limited by several factors such as the frequency of excitation signals, the signal-to-noise ratio, and the pixel size. We propose to use a hypergraphs image inpainting technique for SAM that fills in missing information to improve the resolution of the SAM image. We compared the performance of our technique with four other different techniques based on generative adversarial networks (GANs), including AOTGAN, DeepFill v2, Edge-Connect and DMFN. Our results show that the hypergraphs image inpainting model provides the SOTA average SSIM of 0.82 *with a PSNR of* 27.96 *for* $4\times$ *image size enhancement over the raw SAM image. We emphasize the importance of hypergraphs' interpretability to bridge the gap between human and machine perception, particularly for robust image recovery tools for acoustic scan imaging. We show that combining SAM with hypergraphs can yield more noise-robust explanations.*

1. Introduction

The role of computer-aided diagnosis and application in the current digital revolution cannot be overstated. One such active area of research in computer vision is image inpainting, which involves reconstructing or repairing images while being unnoticed by the casual observer. It is used to fix damaged or corrupted areas of an image, eliminate undesired elements from images, and fill in missing or occluded parts of an image. Object removal, high-resolution imaging, image-based blending, and image denoising are just some of the many uses for patch filling.

The key challenge of image inpainting lies in synthesizing both global semantic visual perceptions and local textured patterns that are coherent with background regions [51]. Deep learning-based methods have addressed the long-standing limitations of traditional inpainting methods by utilizing the information already present in an image to infer missing information. As a result, it has significantly improved the quality of the final output.

Despite the recent success of deep learning in image inpainting for biomedical research, it is still challenging to apply these methods [28, 35, 46] to the limited availability of training data in acoustic microscopy imaging owing to high cost of data acquisition and the need for specialized equipment. Compared to conventional deep learning-based inpainting applications, the complexity of acoustic microscopy images, with variable degrees of contrast and noise, and an uneven distribution of missing data, may cause issues in accurate prediction. They fail to identify the global context or semantics of the image, resulting in implausible results. Real-time non-invasive processing may be required in some acoustic microscopy imaging applications. Finally, deep learning methods frequently behave like a black-box, making it hard to comprehend how they make predictions or detect problems when they fail to deliver accurate results. This can make it difficult to utilize these models in critical settings, such as medical imaging.

1.1. Scanning acoustic microscopy

High-frequency Scanning Acoustic Microscopy (SAM) is a highly sensitive and precise technique for imaging the surface and subsurface structures of various materials. It employs 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.

[*]These authors contributed equally to this work.

SAM is capable of non-invasive micro-structural characterization of different industrial objects and biological specimens [4, 17]. For example, the microelectronics and semiconductor industries are highly competitive and demanding markets. SAM technology is critical in the development of improved mold designs for flip-chip packages in this context. It can be used to characterize and determine the mechanical properties of piezoelectric materials, structural health monitor (SHM) of composite structures, detect surface defects in polymer circuits, and study the propagation of isotropic or anisotropic phonons [13–16, 34, 37]. Furthermore, 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 [41, 42].

The quality of images produced by SAM at a given 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, thereby reducing the scanning time. In contrast, high-resolution images at the same frequency necessitate a greater number of scanning points, resulting in longer data acquisition times. Data acquisition is essential for imaging biological specimens, and high resolution with smaller step sizes is ideal. However, larger step sizes in scanning can result in degraded image quality due to fewer objects' information. To address this issue, conventional image interpolation or learning-based inpainting techniques can be used to improve image quality [3, 30, 33].

The proposed model in this paper draws heavily on research conducted by Wadhwa *et al*. [36] in 2021. To reconstruct missing regions, the model leverages hypergraph structures to identify and incorporate similar features from the background. The model is divided into two stages, coarse and fine, with the goal of producing results that accurately capture the overall context, as well as finer details of the image. The framework includes a trainable method to compute a data-dependent incidence matrix for hypergraph convolutions. The local consistency of the image is ensured by gated convolution rather than regular convolution in the discriminator network.

The following are the primary contributions of this work:

1. As far as we know, there have been no prior reports on image inpainting in acoustic microscopy. We have shown image inpainting for a large collection of SAM images.
2. Our image inpainting strategy uses an alternative hole mask to input the image and generate a high-resolution version using coarse-to-fine hypergraphs strategy.
3. We demonstrate that a two-stage exemplar-guide framework is able to produce higher-quality inpainting results than recent SOTA on challenging datasets, including publicly accessible CelebA-HQ dataset.

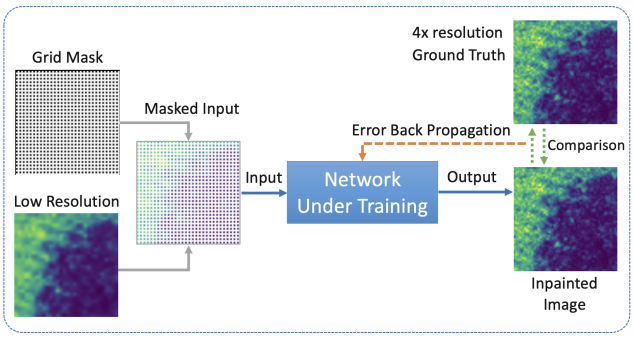

Figure 1. Illustration of the overarching image inpainting strategy for SAM. We utilize the alternative hole mask to provide an input image for the model and then employ image inpainting to produce a high-resolution version of the original.

4. The approach opens up the possibility of removing the barriers of step-size limitations of high-speed imaging for biological samples.

Figure 1 depicts the overall strategy used in this paper.

2. Background

There are two major aspects in generating a contextually plausible and realistic image: (a) global semantic structure and (b) fine detailed texture surrounding the gaps. Image inpainting can accomplish this using (i) content/texture adapting methods and (ii) learning-based methods.

The first method employs simple patch matching algorithms that iteratively fill the missing pixels by searching for similar patches from the neighboring non-missing pixels in the image [2, 7, 10, 11]. Earlier techniques utilized concepts similar to exemplar-based texture synthesis [7], i.e., matching and copying background patches into holes from low-resolution to high-resolution or propagating from hole boundaries. Although these methods are effective in synthesizing a texture-consistent output [39, 49], they are incapable of producing semantically meaningful content. Recently, diffusion-based approaches [1, 27] use variational algorithms or patch similarity to propagate background information into missing regions. Although successful in filling small or narrow regions, these methods struggle when faced with more substantial voids. Unlike diffusion-based approaches, patch-based systems can utilize texture synthesis techniques to fill in large missing regions. Ding *et al*. [9] proposed an exemplar-based method to effectively inpaint geometric patterns and textures. Due to the fact that these approaches primarily utilize low-level features for patch matching, they are incapable of filling in voids with semantically meaningful or novel content.

The second approach, i.e., learning-based techniques employ GANs, an effective approach for generative modeling using deep learning techniques like convolutional neu-

ral networks (CNNs). The GAN framework consists of two main components, a generator that is trained on a dataset that intelligently discovers and learns the features and patterns to generate new examples and a discriminator that distinguishes between the generated examples and the original ones in the dataset. The discriminator helps ensure that the generator produces plausible outputs that fall within the same domain as the original data.

For several years, GAN-based techniques have been employed for image inpainting [5, 23, 32]. The first generation of deep CNNs [43] was optimized to fill small, narrow areas of missing data. Pathak *et al*. [32] were the first to create GAN-based image-inpainting approaches to fill in large holes; their method involved training an encoder-decoder network to deal with holes of size 64×64 and infer semantic content. The core concept of channel-wise fully connected layer served as a baseline for many subsequent models. Iizuka *et al*. [21] takes image inpainting to the next level by including 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. Almost all of the following image-inpainting papers adopted this multi-scale discriminator architecture.

The next big breakthrough came in 2018 with Yu *et al*. [46]'s work DeepFill v1. Using a contextual attention (CA) layer, which is a fully convolutional differential layer that assigns weights to individual features showing their contribution to each location in the missing region, this model outperformed its predecessor, Shift-Net [45]. Yu *et al*. [47] published an improved version of this model in 2019 called DeepFill v2. Gated convolutions are the most crucial part of this paradigm. Later, Zeng *et al*. [50] adopted an end-to-end deep generative model and a nearest neighbor based global matching. This method, however, is primarily trained on a large centering square mask and does not generalize well to masks of arbitrary shape, size, and location. To better manage irregular masks, partial convolution was developed for image inpainting [28], with the convolution being masked and re-normalized to use only valid pixels. [38] used attention mechanism and partial convolution to achieve more realistic inpainting results. The focus has thus far been on applying image inpainting techniques to optical images, and there has been a lack of research on applying these techniques to SAM images.

Furthermore, learning models are oftentimes obscure and hard to interpret, making the inpainting output difficult to control or modify. Many generative models [22, 51] have recently been proposed to handle the problem under circumstances. Lee *et al*. [26] introduced a diversity image-to-image translation method based on disentangled representations, with limited style codes retrieved from an encoder network. This constraint stems from the fact that learning-based generation models inevitably use pre-determined labels, resulting in the same output for each domain.

3. Materials and method

Image inpainting refers to the task of reconstructing missing regions of an image. We aim to use this techniques for image high-resolution. This was achieved through the use of mask fabrication. The mask used to train the dataset is a black matrix for every 3 white pixels. The lower-resolution image is enlarged by 4 times by inserting 3 white pixels between each recorded data point.

Training and directly employing the model for $4\times$ inpainting does not give the SOTA results. As a result, the model was trained to perform a $2\times$ up-sampling by incorporating a 1 white pixel between each recorded data point. The white pixels are then filled with image inpainting techniques. Another subsequent $2\times$ up-sampling is repeated to produce an overall $4\times$ up-sampled result. The following subsection describes the hypergraphs architecture in detail.

3.1. Hypergraphs

The hypergraph is constructed using a combination of spatial and feature-based clustering techniques, which capture both local and global structures in the image. A hypergraph, defined by $G = (V, E, W)$, consists of a set of hyperedges $E = \{e_1, ..., e_n\}$ that connect two or more vertices $V = \{v_1, ..., v_n\}$ 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 $H \in \mathbb{R}^{N\times M}$, where the link is expressed as:

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

Given a hypergrapgh G, vertex degree $D \in \mathbb{R}^{N\times N}$, and hyperedge degree $B \in \mathbb{R}^{M\times M}$ are expressed in Eq. (2).

$$D_{ii} = \sum_{e=1}^{M} W_{ee} H_{ie} \quad ; B_{ee} = \sum_{i=1}^{N} H_{ie}. \tag{2}$$

Equation (3) computes the normalized Hypergraph Laplacian matrix $\Delta \in \mathbb{R}^{N\times N}$, where the matrix is symmetric positive semi-definite [52].

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

Using the eigendecomposition of Δ given by $\Delta = \Phi\Lambda\Phi^T$, we obtain eigenvectors $\Phi = \{\phi_1, ..., \phi_N\}$ and the diagonal matrix $\Lambda = \text{diag}(\lambda_1, ..., \lambda_N)$, which contains the associated non-negative eigenvalues. $\hat{x} = \Phi^T x$ denotes the hypergraph Fourier transform. Equation (4) describe the convolution on the signal $x \in \mathbb{R}^N$, where $g(\Lambda) = diag(g(\lambda_1), ..., g(\lambda_N))$ is a function of Fourier coefficients.

$$g \odot x = \Phi g(\Lambda)\Phi_T x \tag{4}$$

The convolution operation on the hypergraph signal can be described by parameterizing $g(\Lambda)$ using truncated Chebyshev polynomials up to K^{th} order [8].

$$g \odot x = \sum_{k=0}^{K} \theta_k T_k(\Delta) x \tag{5}$$

$$g \odot x = \theta D^{-\frac{1}{2}} H W B^{-1} H^T D^{-\frac{1}{2}} x \tag{6}$$

The convolution process can be generalized to the multi-layer hypergraph convolution network in Eq. (7) for a given hypergraph signal $X^l \in \mathbb{R}^{N \times C_t}$,

$$X^{l+1} = \sigma(D^{-\frac{1}{2}} H W B^{-1} H^T D^{-\frac{1}{2}} X^l \Theta) \tag{7}$$

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

3.2. Hypergraphs convolution on spatial features

Simple graphs can be considered as a special case of hypergraphs in which each hyperedge connects 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, each spatial feature is considered as a node with dimension c, having 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 [12, 44], cross-correlation of the spatial features are used to calculate the contribution of each node to the hyperedge. As a result,

$$H = \Psi(X)\Lambda(X)\Psi(X)^T\Omega(X) \tag{8}$$

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 vector of features after the linear 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 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 stated in [18], and $\Omega(X)$ is implemented using the 7×7 filter. Thus, we arrive at:

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

$$\Psi(X^l) = \text{conv}(X^l, W_\Psi^l) \tag{10}$$

$$\Lambda(X^l) = \text{diag}(\text{conv}(\hat{X}^l, W_\Lambda^l)) \tag{11}$$

$$\Omega(X^l) = \text{conv}(X^l, W_\Omega^l) \tag{12}$$

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 $W_\Psi^l, W_\Lambda^l, W_\Omega^l$ are the learnable parameters for linear embedding. Absolute values are used in the incident matrix to avoid imaginary values in the degree matrices. Hence, hypergraph convolution layer on spatial features can be written as,

$$X^{l+1} = \sigma(\Delta X^l \Theta) \tag{13}$$

where $\Theta \in \mathbb{R}^{C_l \times C_{l+1}}$ is the learnable parameter and σ is the ELU [6] non-linear activation function.

3.3. Architecture and training parameters

Figure 2 depicts the architecture of the hypergraphs image inpainting model. It consists of a two-stage course-to-fine network architecture. While the course network roughly fills the missing region, which is naively blended with the input image, the refined 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 of the image. Dilated convolutions [21] are used to expand the coarse further and refine networks' receptive field. Also, gated convolutions [48] are used to improve performance which can be defined as:

$$Gating = \text{conv}(W_g, I) \tag{14}$$

$$Features = \text{conv}(W_f, I) \tag{15}$$

$$O = \phi(Features) \odot \sigma(Gating) \tag{16}$$

where W_g and W_f are two different learnable parameters for convolution operation, σ is the sigmoid activation function, and ϕ is a non-linear activation function, such as ReLU, ELU, and LeakyReLU. Also, to prevent deterioration of the color coherence of the completed image, batch normalization is removed [21]. In our method, the discriminator has an architecture similar to PatchGAN [22]. All convolution layers are replaced with a gated convolution using which enforces local consistency in the completed image. The discriminator is provided with both 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,

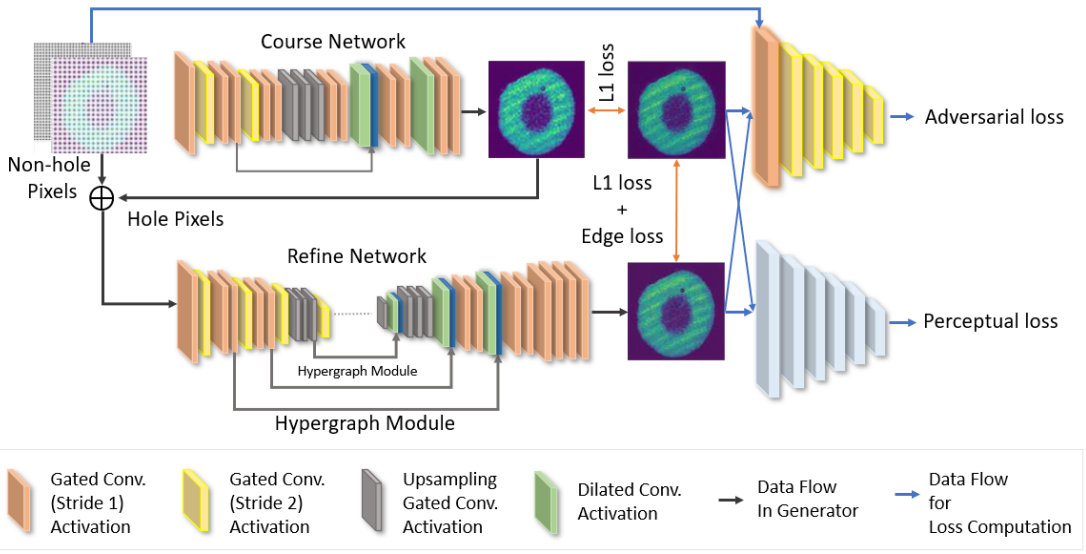

Figure 2. Two-stage course-to-fine hypergraphs image inpainting model architecture. The course network approximately fills the missing region, which is naively blended with the input image, while the refined network predicts finer results with sharp edges.

and edge loss. $L1$ loss is used on both I_{coarse} and I_{refine} to maintain pixel level consistency. Hence, the content loss is defined as,

$$L_{\text{hole}} = ||R\odot(I_{\text{refine}}-I_{\text{gt}})||_1+\frac{1}{2}||R\odot(I_{\text{coarse}}-I_{\text{gt}})||_1 \quad (17)$$

$$\begin{aligned} L_{\text{valid}} =& ||(1-R)\odot(I_{\text{refine}}-I_{\text{gt}})||_1 \\ &+\frac{1}{2}||(1-R)\odot(I_{\text{coarse}}-I_{\text{gt}})||_1 \end{aligned} \quad (18)$$

where L_{hole} is the loss for the hole pixel values, and L_{valid} is the loss for the non-pixel values.

For a given input x, let $\phi_l(x)$ denote the high-dimension features of the l^{th} activation layer of the pre-trained network, then perpetual loss is defined as,

$$L_p = \sum_l ||\phi_l(G(I_{\text{im}})) - \phi_l(I_{\text{gt}})||_1 \quad (19)$$

The perceptual loss for final prediction I_{refine} and I_{comp} is computed, where I_{comp} is the final prediction, but the non-hole pixels are set directly to ground truth [22]. Edge-preserving loss [31] is used to maintain edges in the predicted images, which can be defined as,

$$L_{\text{edge}} = ||E(I_{\text{refine}}) - E(I_{\text{gt}})||_1 \quad (20)$$

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

$$L_{\text{total}} = \lambda_{\text{hole}}L_{\text{hole}}+\lambda_{\text{valid}}L_{\text{valid}}+\lambda_{\text{adv}}L_{\text{adv}}+\lambda_p L_p+\lambda_{\text{edge}}L_{\text{edge}} \quad (21)$$

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

3.4. Experimental dataset

We used a high-resolution scan in SAM to create $4\times$ high-resolution images. A total of 33 high-resolution images were 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 of dimension 96×96 pixels to create diversity, and making the network training robust. Cropping is done by starting from the image's top-left corner and striding in the G-direction and H-direction. This was done to maintain a uniform size for training and to ensure that the images' overall semantics is somewhat preserved.

The training data set consisted of 402 such 96×96 images scaled in the range [0-1] during training. For each high resolution crop, a corresponding low resolution crop was created by masking 3 pixels in a 2×2 window with a stride size 2, and retaining only the top-left pixel in each of these windows. The mask is inspired by the fact that the scanning acoustic microscope is operated on two different step sizes. The low-resolution images are recorded as having a step size of $200\mu m$ in contrast to the $50\mu m$ step size of the higher-resolution images used for model training.

3.5. Experimental setup

The SAM has two operational modes: reflection and transmission mode. Figure 3 presents a labeled image of SAM that is utilized for image acquisition. Further details regarding the working principles of these modes can be found elsewhere. In this paper, we have focused on the reflection mode to scan the samples. To focus acous-

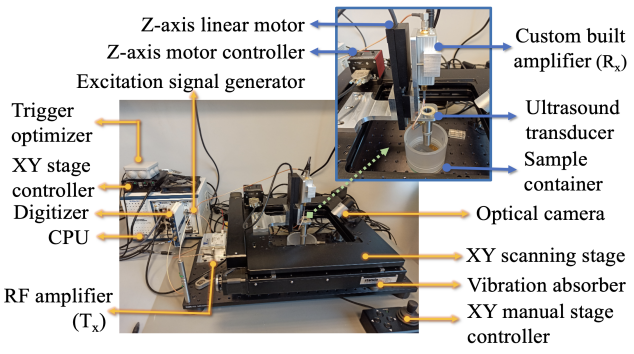

Figure 3. A tagged image of SAM, which is used for image acquisition. The experimental setup displays all of the key elements.

tic 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 was repeated at various points in the XY plane. Alternatively, a C-scan can be viewed as the summation of A-scans in two dimensions.

Experimental data were collected using a custom-built SAM (shown in Fig. 3) that included a high-precision scanning stage from Standa (8MTF-200-Motorized XY Microscope Stage) and controlled by a LabVIEW program. Previous work by Kumar *et al*. [25] utilized a similar experimental setup to correct for inclined samples. 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). 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 $6.35mm$ aperture and a $12mm$ 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 $50\mu m$ steps. Low-resolution images were obtained using a 20 MHz transducer with a focal length of $50mm$. All experiments were carried out in distilled water while maintaining a constant room temperature of approximately 22 °C. To evaluate the models, a discarded reindeer antler was used as a biological sample for imaging. Prior to scanning, the moss on the antler was removed by cleaning it with lukewarm water and 96% ethanol. The sample was then 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.

We use two experiments to analyze the inpainting performance of the proposed approach. First, we study the effect of various generative learning-based models on inpainting outputs. Second, we compare our final outputs from our hypergraph method with outputs from SOTA non-learning-based inpainting techniques to evaluate the high-frequency information compensation and compare global matching with compositional matching. Later, we evaluated the diversity of our final results by varying the exemplars.

4. Results and discussion

Four image inpainting techniques, namely AOT-GAN [11], DeepFill v2 [48], Edge-Connect [29] and DMFN [20] were compared with the hypergraphs image inpainting [36]. These are tested on the CelebA-HQ dataset [24]as well as the SAM dataset. The results are discussed below.

4.1. Results on the CelebA-HQ dataset

Figure 4 shows the outputs of some of the CelebA-HQ dataset along with individual metric evaluation. A PSNR score (labeled in yellow) of 23.80 and an SSIM score (labeled in white) of 0.74 were obtained for the first image sampled. The mean SSIM and PSNR scores of all the models over the CelebA-HQ dataset are shown in Tab. 1. Quantitative comparison of five sampled images in Figs. 5 and 6 clearly validates the superiority of hypergraphs used with our masked input over other methods.

The SOTA's poor performance is due to the suggested application employing the SAM simulated grid-mask to im-

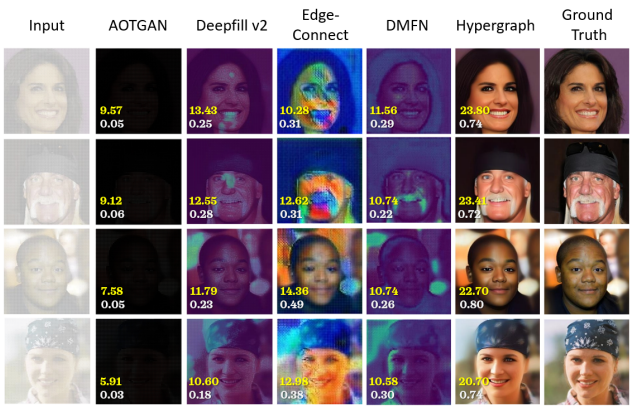

Figure 4. Qualitative comparision of models output on CelebA-HQ dataset. (Parts of work reproduced under CC-BY-4.0) [19,24].

Model	AOT-GAN	DeepFill v2	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 1. Quantitative comparison of average SSIM & PSNR scores for all comparing models tested on the CelebA-HQ datset.

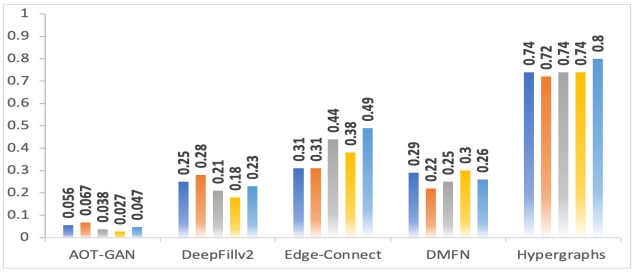

Figure 5. SSIM scores comparision on 5 randomly sampled CelebA-HQ dataset images.

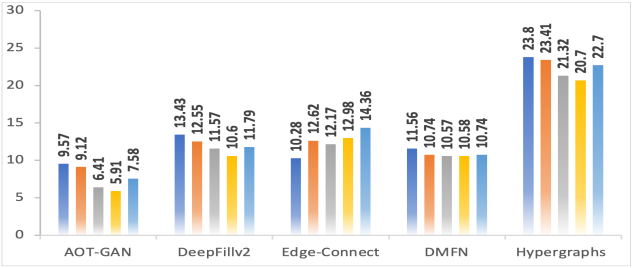

Figure 6. PSNR scores comparision for the methods on 5 sampled CelebA-HQ dataset images.

paint 75-90% of the missing region for 4x resolution output. Wheras, the conventional SOTA model training shape mask only impaints 20-25% of the missing regular/irregular shaped area and yields better results. We used the CelebA-HQ dataset as a baseline to show how other inpainting methods fail to leverage SAM using grid-mask.

4.2. Results on SAM images dataset

The average SSIM and PSNR scores of all five generative learning models on the SAM testing dataset are reported in Tab. 2. Of all popular approaches, the hypergraph approach produced the SOTA result, as evident by the metric score in the table, as well as visual inspection of Figs. 7 and 8. The average PSNR score of 27.96 ± 2.98 and the SSIM score of 0.8234 ± 0.10 were obtained for 50 test set images using the hypergraphs model. We compared the result to classical bilinear interpolation and ESRGAN [40], a popular super-resolution technique. ESRGAN ranked second in our evaluations, however, it underperformed hypergraphs in terms of both SSIM and PSNR. The individual PSNR and SSIM values (y-axis) for the entire test set are shown against the DeepFill v2 metric value (x-axis) in Fig. 7. The advantage of the suggested method is clearly projected by the linear fit regression line for the hypergraph (marked in yellow). We found that hypergraphs were the most effective model for learning to restore the missing pixels in our mask, resulting nearly four times larger images.

An abalation study on the network parameters for the SAM dataset by systematically removing the gated convolution, resulted in a 3 ± 0.52 decrease in PSNR and a 0.1 ± 0.04 decrease in SSIM value. Individually, we replaced the ELU activation unit in gated convolution with a LeakyReLU, implemented non-incremental learning, and added a trainable VGG19 model for the perceptual loss. All of these factors had a negative impact on the network for this application, resulting in an average decline in both PSNR and SSIM.

Using the visual turing test (VTT), we observe that the generated images are better than the Celeba-HQ dataset results (ref. Fig. 4). In contrast to the CelebA-HQ dataset, the resulting images here are more evenly distributed throughout the models. We suspect that this is due to the fact that all of the images in our dataset share a large array of features. However, a deeper VTT inspection will show that the hypergraphs model did get the closest findings to the ground truth when compared to the other methods.

We also compared the results obtained in this paper to those obtained using more conventional digital resolution enhancement techniques, such as bilinear, Lanczos, and nearest neighbor methods. The results of the hypergraph model were shown to be comparable to those of more conventional heuristic-based strategies. Nevertheless, it should be noted that the primary motivation for the effort was to evaluate our novel grid mask, which can simulate super-resolution, against AI-based inpainting models.

	DeepFill v2	AOT-GAN	Edge-Connect	DMFN	Hypergraph
PSNR	18.09 ± 3.29	22.47 ± 2.05	20.64 ± 2.80	19.46 ± 3.89	27.96 ± 2.98
SSIM	0.43 ± 0.21	0.55 ± 0.13	0.54 ± 0.18	0.50 ± 0.17	0.82 ± 0.10

Table 2. Average SSIM & PSNR scores of various generative learning-based models on our acoustic scan testing dataset.

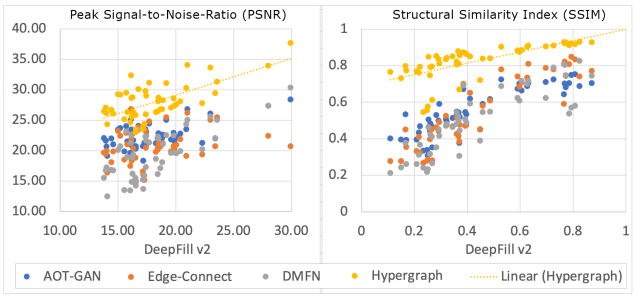

Figure 7. Scatter plot for image similarity metrics to represent the trendline of competing models compared to DeepFill v2 as the baseline performing model.

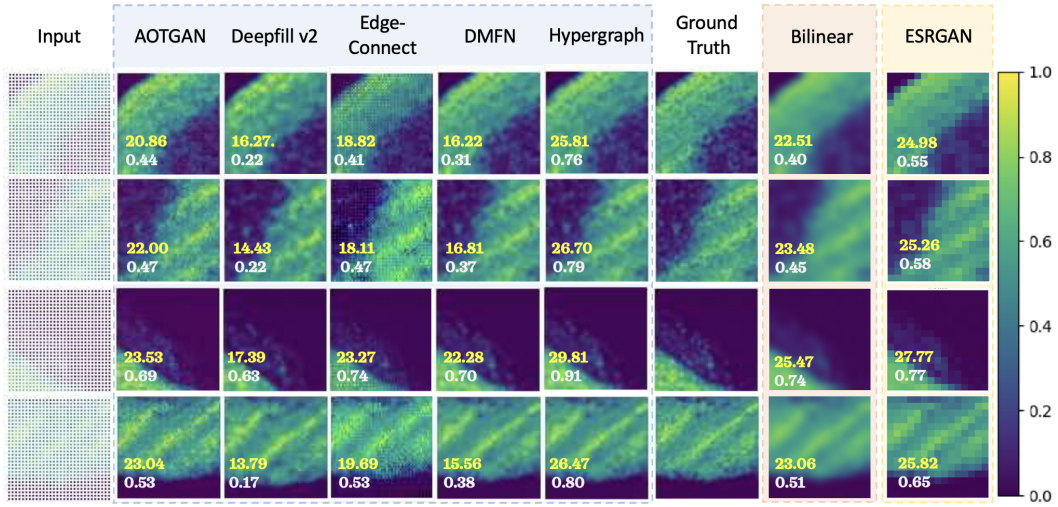

Figure 8. Randomly sampled result for inpainting task on our acoustic scan dataset. From left to right, the corresponding outputs of various models using our super-resolution method.

4.3. Interpretability implications of the proposed approach

Interpretable Deep Learning (IDL) is a field that focuses on developing DL models that are transparent and can be understood by humans. Generally speaking, imparting interpretability in image inpainting helps to shed light into model's prediction and reveals how the model fills in the missing pixels. This study addresses the challenges of using deep generative models to inpaint large spaces and presents a hypergraph-based approach to improve interpretability and transparency.

Oftentimes, we are mesmerized by the convincing image similarity metrics like PSNR, SSIM and not by their effectiveness in achieving intended results. It is possible that the high-quality image generated by these techniques is contextually irrelevant to the task at hand. To test the efficacy of the proposed hypergraph-based inpainting technique, we create our own acoustic imaging scan dataset in this paper. The generated hypergraph is then used to direct the inpainting process by specifying the connections between pixels and features that should be preserved in the output image.

The use of hypergraph-based methods for image inpainting offers the advantage of incorporating prior knowledge or constraints into the process, which can help preserve known features or structures in the image. Additionally, these methods can provide transparency and interpretability through visualization techniques such as node-link diagrams and heatmap visualizations and graph-theory metrics such as centrality or clustering coefficient for the hypergraph's nodes and hyperedges. These techniques can aid in gaining insight into the structure of the hypergraph and the inpainting process, which can be useful for further refinement of the technique. Further research in this field is needed to overcome large-step acoustic imaging with robust generalizable solutions.

5. Conclusion

In this study, we present the development of a deep learning-assisted acoustic microscopy system to improve the image resolution of industrial and biological samples through image inpainting. We start with creating a mask of alternate data points and white pixels and then used subsequent repetitive image inpainting to $4\times$ upsample the image resolution. Five popular learning-based methods were employed to fill the missing pixels in the SAM images, namely AOT GAN, DeepFill v2, Edge-Connect, DMFN, and Hypergraphs. The idea of comparing the inpainting method to a super-resolution problem rests on the fact that both can be interpreted in different ways by diverse people. The hypergraphs model presented the SOTA results in terms of SSIM and PSNR for both our acoustic scan dataset and CelebA-HQ data. The hypergraphs image inpainting network consists of a two-stage course-to-fine network architecture. The refine network predicts sharper outcomes with sharp edges, whereas the course network loosely fills the missing region, which is then naively blended with the input image. Transfer learning was considered in the process to prevent the model from overfitting, as limited (800) images were used for training purposes.

6. Acknowledgement

This work was supported by the Research Council of Norway Project No. 325741 and VirtualStain(UiT) Cristin Project ID: 2061348.

References

[1] Coloma Ballester, Marcelo Bertalmio, Vicent Caselles, Guillermo Sapiro, and Joan Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE transactions on image processing*, 10(8):1200–1211, 2001. 2

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

[3] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, 2000. 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. 3

[6] 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. 4

[7] 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

[8] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016. 4

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

[10] Iddo Drori, Daniel Cohen-Or, and Hezy Yeshurun. Fragment-based image completion. In *ACM SIGGRAPH 2003 Papers*, pages 303–312. 2003. 2

[11] 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. IEEE, 1999. 2, 6

[12] 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

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

[14] A Habib, A Shelke, M Vogel, S Brand, Xin Jiang, U Pietsch, S 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. 2

[15] A Habib, A Shelke, M Vogel, U Pietsch, Xin Jiang, and T Kundu. Mechanical characterization of sintered piezoelectric ceramic material using scanning acoustic microscope. *Ultrasonics*, 52(8):989–995, 2012. 2

[16] 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 (IUS)*, pages 1–4. IEEE, 2018. 2

[17] 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

[18] 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. 4

[19] Huaibo Huang, Ran He, Zhenan Sun, Tieniu Tan, et al. Introvae: Introspective variational autoencoders for photographic image synthesis. *Advances in neural information processing systems*, 31, 2018. 6

[20] Zheng Hui, Jie Li, Xiumei Wang, and Xinbo Gao. Image fine-grained inpainting. *arXiv:2002.02609*, 2020. 6

[21] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017. 3, 4

[22] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 3, 4, 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. 3

[24] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. 6

[25] 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

[26] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*, pages 35–51, 2018. 3

[27] Anat Levin, Assaf Zomet, and Yair Weiss. Learning how to inpaint from global image statistics. In *ICCV*, volume 1, pages 305–312, 2003. 2

[28] 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 (ECCV)*, pages 85–100, 2018. 1, 3

[29] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019. 6

[30] Mark Nitzberg, David Mumford, and Takahiro Shiota. *Filtering, segmentation and depth*, volume 662. Springer, 1993. 2

[31] Ram Krishna Pandey, Nabagata Saha, Samarjit Karmakar, and AG Ramakrishnan. Msce: An edge-preserving robust loss function for improving super-resolution algorithms. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part VI 25*, pages 566–575. Springer, 2018. 5

[32] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 3

[33] Zhen Qin, Qingliang Zeng, Yixin Zong, and Fan Xu. Image inpainting based on deep learning: A review. *Displays*, 69:102028, 2021. 2

[34] BR 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*, pages 423–449e. Elsevier, 2013. 2

[35] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1

[36] 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, 6

[37] Sanat Wagle, Anowarul Habib, and Frank Melandsø. Ultrasonic measurements of surface defects on flexible circuits using high-frequency focused polymer transducers. *Japanese Journal of Applied Physics*, 56(7S1):07JC05, 2017. 2

[38] Ning Wang, Sihan Ma, Jingyuan Li, Yipeng Zhang, and Lefei Zhang. Multistage attention network for image inpainting. *Pattern Recognition*, 106:107448, 2020. 3

[39] Shuenn-Shyang Wang and Sung-Lin Tsai. Automatic image authentication and recovery using fractal code embedding and image inpainting. *Pattern Recognition*, 41(2):701–712, 2008. 2

[40] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 7

[41] Mario Wolf, Peter Hoffrogge, Elfgard Kühnicke, Peter Czurratis, and Christian Kupsch. Inspection of multilayered electronic devices via scanning acoustic microscopy using synthetic aperture focusing technique. In *2022 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2022. 2

[42] M Wolf, A Sukumaran Nair, P Hoffrogge, E Kühnicke, and P Czurratis. Improved failure analysis in scanning acoustic microscopy via advanced signal processing techniques. *Microelectronics Reliability*, 138:114618, 2022. 2

[43] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. *Advances in neural information processing systems*, 25, 2012. 3

[44] 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

[45] Zhaoyi Yan, Xiaoming Li, Mu Li, Wangmeng Zuo, and Shiguang Shan. Shift-net: Image inpainting via deep feature rearrangement. In *Proceedings of the European conference on computer vision (ECCV)*, pages 1–17, 2018. 3

[46] 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. 1, 3

[47] 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. 3

[48] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 4, 6

[49] 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 (DSP)*, pages 1–5. IEEE, 2018. 2

[50] Yuan Zeng, Yi Gong, and Xiangrui Zeng. Controllable digital restoration of ancient paintings using convolutional neural network and nearest neighbor. *Pattern Recognition Letters*, 133:158–164, 2020. 3

[51] Yuan Zeng, Yi Gong, and Jin Zhang. Feature learning and patch matching for diverse image inpainting. *Pattern Recognition*, 119:108036, 2021. 1, 3

[52] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. *Advances in neural information processing systems*, 19, 2006. 3