Advances in Computer, Signals and Systems (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2022.060305
ISSN 2371-8838 Vol. 6 Num. 3

Phase Completion for Fringe Projection Profiler Based on Neural Networks

Ziyu Yin[1,a,*], Junzheng Li[2,b]

[1]*Department of Game, Software Engineering Institute of Guangzhou, Guangzhou, China*
[2]*School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China*
[a]*yzy@mail.seig.edu.cn*, [b]*lijunzheng@sjtu.edu.cn*
**Corresponding author*

Keywords: fringe projection profiler (FPP), physical rendering, neural network, optical measurement

Abstract: Fringe projection profiler (FPP) measures the geometry of the target surface by projecting the pre-modulated stripe map onto the surface, and then capture the phase map with a camera. However, the inaccurate exposure or the characteristics of the surface reflectance may influence the imaging quality of the phase map, leaving some over-exposure and under-exposure regions. Addressing to this problem, this paper propose to apply a neural network to complete the phase map. Firstly, we propose a synthetic dataset to simulate the phase map of the inaccurate exposure regions, based on a physical rendering model. After that, we implement a transformer neural network to complete the missing phase information. Experiments show that the proposed neural network can complete the missing information from its neighbouring information, and provide precise completion results.

1. Introduction

Fringe projection profiler (FPP) is a measurement approach that has been widely used in the industry [1] due to its accuracy and efficiency [2]. A typical framework of FPP is shown in Figure 1. The projector projects the pre-modulated pattern onto the target surface. After that, the stripe pattern (or the phase map) is presented on the surface. Through capturing the phase map with a camera and analysis it with a reconstruction algorithm, the depth map of the surface is obtained.

Although being widely used in the industry, the precision and robustness of the FPP system is strongly affected by the surface characteristics, such as the material, the curvature and the reflectance. Especially for those surface with complex geometry and made of metal, as shown in Figure 1, the phase map may have a wide dynamic range. However, the dynamic range of a camera when specific parameters are given, is relatively low. Therefore, over-exposure and under-exposure happens in some regions, leaving missed information for the depth reconstruction algorithms.

The community has recently introduced multi-exposure [3] and adaptive projection [5], [6] to improve the dynamic range of the camera, from the perspective of setting different projector and camera parameters and capture a group of images. Furthermore, these schemes extract the well-measured features from different images under different projection-camera conditions. However, these schemes could be complex for implementation, and take significantly longer time for each

measurement instance.

On the other hand, we have noticed the impressive achievements that deep learning has made [7] in areas such as computer vision and decision making. Specially, deep learning has made great success in image completion and produces realistic images. For example, Xie et al. [8] proposed an auto-encoder architecture based on CNN for image completion. Furthermore, transformer architectures are introduced to improve the high-level feature representation ability [9].

This paper focuses to complete over-exposure regions and forms reliable results based on the deep learning approaches and avoiding the multi-exposure or adaptive-projection pipelines. Firstly, we propose a physical model which based rendering algorithm to generate a synthetic dataset for network training. After that, we apply the advanced image completion neural network to complete the over-exposure regions based on the neighbour features. Finally, this pipeline is tested in experiments to validate its effectiveness.

The paper is organized as follows. The second section reviews the advanced methods in projection methods and image completion neural networks. The third section details the dataset and completion model. The fourth section provides experiments to validate the proposed method. And the last section concludes this paper.

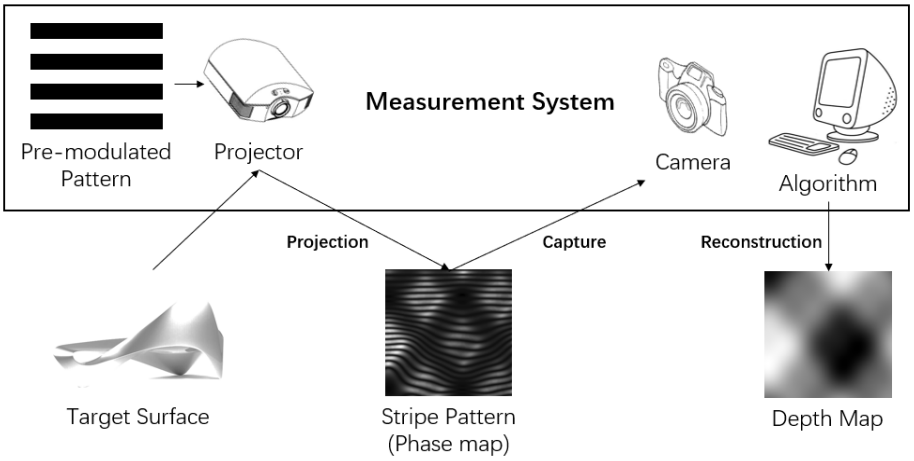

Figure 1: Framework of a fringe projection profiler measurement system

2. Related works

2.1. Advance Projection Methods

The robustness of FPP system is strongly affected by the surface geometry and reflectance. When a fixed pre-modulated pattern is projected onto the target surface, the dynamic range of the phase map could significantly exceed the dynamic range of a camera. Therefore, the community proposes advanced projection methods, including the multi-exposure and adaptive-projection to eliminate this problem.

Multi-exposure is proposed to capture the information in the shiny regions [3]. This pipeline sets different projection intensity and exposure time to balance the over-exposure regions and the under-exposure regions, capturing a series of images [6]. After that, a fusion algorithm is proposed to synthesis a single phase map from the raw images. And finally the 3D geometry of the target surface can be calculated accordingly. Furthermore, Tang et al. [4] proposes an optimization algorithm to figure out the best frequency and phase shift based on this pipeline.

Adaptive projection is another method to eliminate the over-exposure problem. This method focus to optimize the pre-modulated pattern though adjusting the local intensity [5]. For example, for those

regions that tend to be over-exposure, the local intensity decreases in these regions, and vise versa. Li et al. [10] proposes an adaptive fringe-pattern projection (AFPP) framework following this pipeline. The framework includes three steps: determination of saturated camera pixels, determination of local MIGL adaptation, and generate the adaptive projection pattern. In the measurement pipeline, this pattern is projected onto the target surface. The phase map is captured and the depth map is reconstructed.

Although the advanced projection methods can improve the image quality significantly, they follow the trail-and-adjust fashion that requires well-experienced technicians and complex algorithms. Therefore, we aim to design a fast and precise measurement procedure based on neural networks.

2.2. Image Completion

Image completion task which is also called image inpainting task has become an important research topic in computer vision. Image inpainting is a task that aims to fill the missing pixels or regions of images. Early works are based on diffusion-based methods [11]. They calculate the value of the missing pixels using their undamaged neighbour. For example, patch-based methods calculate the similarity of the patches using the hand-crafted distance metrics and use the patched to fill the missing regions. What's more, another traditional method like partial differential equation, interpolation methods and image statistics are also used in image inpainting. Interpolation methods are widely used in phase completion in the industry. However, traditional methods cannot extract high-level information from the images and thus the results of those methods are not robust enough.

Recently, deep learning has achieved great success on the computer vision. Thanks to the development of GANs and VAEs, CNN-based methods has been widely used in image completion. Those methods generate semantic content of the missing regions by propagating the undamaged pixels. [12] first proposed context encoders for image inpainting based on the encoder-decoder architecture combining with GAN. Afterwards, more sophisticated networks like U-Net architecture based methods has been used in image inpainting. One common concern for generating high-quality is the ability of the network to aggregate local and global feature of the context. Therefore, dilated convolution was used to predict the missing regions. What's more, global and local discrimination was proposed to maintain the consistency of the local and global regions separately. [13] proposed contextual attention layer to extract feature from spatially distant undamaged regions. To deal with irregular masks with any shape, some work proposed to use partial convolutions [14] and gate convolutions [15] rather than use a standard convolution network. In addition to one-stage architecture, multi-stage generation was proposed to complete the images from coarse to fine. In those work, the first-stage network is responsible for generating coarse global content in the missing regions. The second-stage network then generate the local detailed structure based on the coarse global content. In general, CNN-based approaches fill the missing regions from their undamaged neighbouring visible pixels, which are limited by the locality of the convolutional feature extraction mechanism. In phase completion for FPP, the incomplete regions are large and continuous. Therefore, CNN-based methods may have limitations in FPP completion.

Motivated by the great success of transformer architecture in natural language processing(NLP), many researchers applied transformer-based network in computer vision tasks recently. In image inpainting, Deng et al. proposed Contextual Transformer Network for improving the continuity of context. [16] proposed an decoder-encoder based transformer which name masked autoencoder. What's more, [9] develops an inpainting transformer for completing large missing regions, which is the state-of-the-art method up to now.

3. Methods

3.1. Dataset and Rendering Model

Training neural networks requires a significantly large amount of training data while collecting enough paired and labelled data is difficult in real scenes. Therefore, we propose a FPP saturation-exposure dataset based on physical rendering.

We firstly set the simulation environment by giving the parameters such as the position of the projector and the camera, the light axis, and the pre-defined projection pattern, as shown in Figure 2(a). For a certain depth map, we calculate the geometry stripe presented at the camera's sensor according to the shape of the target surface and the environment settings. After that, we introduce the physical model of rendering, i.e. the Lambertian reflection model and modulate the intensity of the stripe. Finally, we recognize the saturation regions and suppress them to the largest possible values.

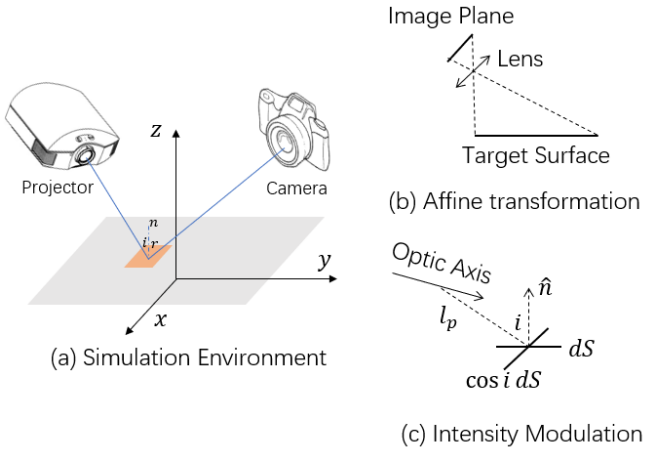

Figure 2: Simulation environment and light path diagram

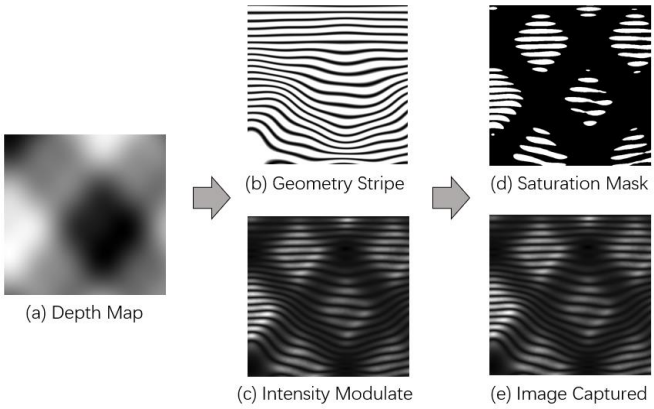

Figure 3: Synthetic process and the dataset

3.1.1. Stripe generation with affine transformation

When projecting the equally spaced stripes onto the target surface, the image captured by the camera is no longer equally spaced. This can be traced to the affine transformation during the imaging process and the geometric shape of the target surface. For most FPP systems, sinusoidal stripes are widely used. Therefore, we firstly set a equally spaced sinusoidal stripes on the object plane. Then we examine each point on the surface and distinguish which part of the stripes it belongs, through the affine transformation, as shown in Figure 2(b). Then we can obtain the pattern on the target surface.

Similar procedures are taken to get the pattern on the camera's plane, as shown in Figure 3(b).

3.1.2. Intensity modulation with Lambertian model

The image in Figure 3(b) seems not realistic. The main reason is that the above method only model the geometric light path without considering the intensity loss during the transmission. Therefore, we borrow the Lambertian reflection model to generate more realistic dataset.

Lambertian model describes the reflection features of diffuse surfaces. It assumes that the reflection intensity is isotropic. The formula of Lambertian model is shown as follow:

$$f(\omega_r, \omega_i) = \frac{dL_r}{dE_i} = \frac{\rho}{\pi} \tag{1}$$

where ω_r, ω_i are the reflection angle and the incidence angle, $f(.)$is the BRDF function, dL_r is the emissivity of reflection, dE_i is the irradiance of incident light, ρ is a constant related to the surface characteristics.

Then we analysis the dE_i component, indicating to what extent the surface dS is lighted by the projector. We assume the projector is the only light source and the differential notation is not necessary.

The main task is to find the stereo corner of dS with respect to the projector. As shown in Figure 2(c), it can be expressed as the follow equation:

$$\mathrm{d}\Omega = \frac{\mathrm{dScos\ i}}{\mathrm{lp}^2} \tag{2}$$

In order to solve the light intensity loss during the transmission, we consider the projector projects isotropic light, which means each unit stereo corner shares the same intensity. Therefore, $E_i \propto d\Omega$.

Similar procedures can be taken for the reflection light. After calculating the intensity loss of each light path, we modulate the geometry stripe Figure 3(b) by multiplication and get the synthetic phase map Figure 3(c).

3.1.3. Saturation suppression and dataset generation

The pattern generated as Figure 3(c) has higher dynamic range than the camera. Therefore, the camera cannot capture the same image as Figure 3(c). We apply the saturation suppression to simulate the over-exposure of the camera.

We set a threshold of the CCD unit in camera, then check each pixel in Figure 3(c) if it exceed the threshold. If it does, we record this pixel in the mask and suppress its value as the max threshold, and finally we get the training set Figure 3(d, e).

During the generation of the dataset, we set sinusoidal surfaces as the depth maps. The depth map is the sum of two sinusoidal functions whose frequencies have 10 times difference.

4. Experiment

4.1. Experiment Settings

We use our render model to generate the phase dataset as our train dataset. And we randomly mask the phase images with a intensity threshold. Our dataset contains more than 5000 paired incomplete-complete synthetic phase images. Nearly 10 % of them are used as the test dataset and another 10% for validation dataset. We train our network using Adam optimizer and we set the learning rate as10^{-4} with decay rate 0.9. In addition, the batch size is set to 8. We use Pytorch as our framework to train the model. And we train our network in 4 Nvidia A30 GPU with more than 200 epochs.

4.2. Completion Model

Given an incomplete phase image, completion model aims to fill the incomplete regions with semantically accurate content. We designed a transformer based neural network to achieve phase completion for fringe projection profiler. Our network is inspired by [16] and [9], which is the simplification of [9] due to memory cost and forward speed considerations. The model has three parts: a convolution-based encoder, a transformer-based module with contextual attention, and a feed-forward reconstruction module.

Firstly, the model takes the incomplete phase image as input and gets the mask image by detecting the overexposure pixels with a threshold filter. Then, Convolution-based encoder is used to extract feature maps from the incomplete phase image and mask image. The convolution-based encoder contains three convolutional layers to down sample the inputs into 1/8 sized feature maps with 180 channels. The feature maps are used as the tokens for transformer-based module. We then borrow the design of transformer-based module from [9] to process the feature map. The transformer-based module employs shifted windows operation guided by dynamical masks to achieve efficient self-attention. This attention mechanism can help the network only pay attention to the valid pixels. The transformer-based module contains three transformer blocks and each block consists of three parts: a self-attention module guide by dynamic masks, a fully connected network and a multilayer perceptron. The last component is a feed-forward reconstruction module. This module aims to generate the complete phase image from the output of transformer-based module. We utilize a three layers convolutional layers to up sample the feature size into the original input image size.

Since the network need to obtain accurate and reasonable results. At the training stage, we use the pixel-wise mean-square-error (MSE) as the loss function. Suppose the ground truth of the phase image is denoted as $gt \in R^{w \times h}$, and the output of the neural network is denoted as $pred \in R^{w \times h}$, where w and h are the scale of the depth map. The loss function is expressed as the following equation:

$$loss = \sum_{i<w, j<h} \left(gt_{i,j} - pred_{i,j}\right)^2 \tag{3}$$

4.3. Phase map completion

To validate our method, we compared our network with interpolation methods, including nearest interpolation and high-performance cubic interpolation. The cubic interpolation uses Qhull to triangulate the input data and construct Bezier polynomial on the triangle. Those interpolation methods are the most commonly used methods in phase map completion task in industry.

To evaluate the methods mentioned above, we use Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) and mean-absolute-error (MAE) as the metric for comparation. The larger the PSNR and SSIM are, and the smaller the MAE is, indicating that the images are more similar to GT and the quality of the images are better. We complete the phase image in the test dataset using the above methods respectively and calculate their PSNR, SSIM and MAE. The results are shown in Table 1.

It can be found that nearest interpolation method shows worst results in all metrics. What's more, cubic interpolation method, as a conventional and high performance method, performs much better than nearest method. Our phase image completion network shows significant higher performance in all metrics, especially in PSNR metric and MAE metric. In SSIM metric, both cubic method and our network are very close to 1, which means that the generation images of both two methods are similar to the GT phase images. However, our method has 28% higher than cubic method in terms of PSNR

and 63% lower than cubic method in terms of MAE. It shows that our deep learning based method can complete the damage phase images with higher quality and less error.

Table 1: Statistics of the inpainting result

Method	PSNR	SSIM	MAE
Nearest	24.0487	0.933	0.0207
Cubic	36.370	0.985	0.0041
Ours	**46.041**	**0.989**	**0.0015**

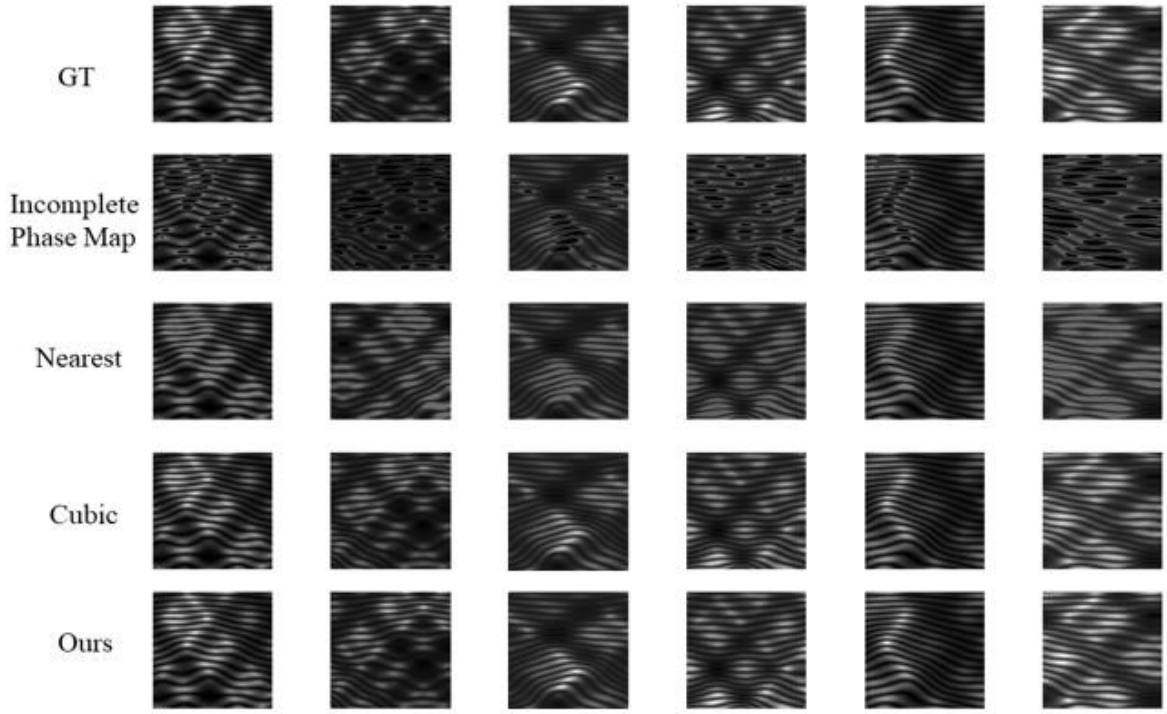

Figure 4: Visualization of the inpainting results among different methods

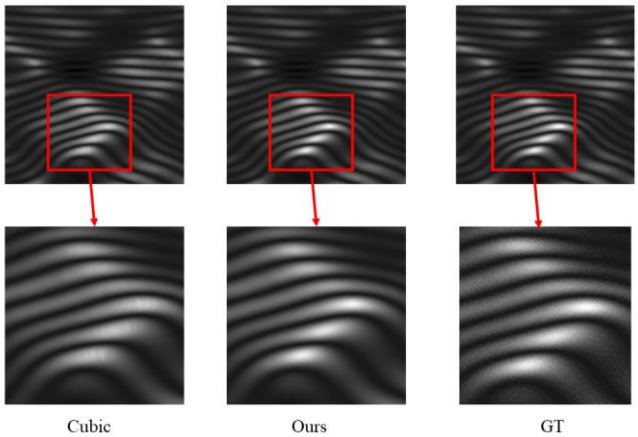

Figure 5: Partial enlargements of the results

In addition to the analysis the metric, we visualize the results of different methods in Figure 4. We choose different masked scales phase images as the samples. When focusing the completed results in the missing regions, the nearest method significant underestimate the phase map at these regions. This is due to the fact that nearest method only uses neighbouring pixels for prediction. Figure 5 shows partial enlargements of cubic method and our methods. Compared to the GT phase image, our

method has the same smooth and consistent intensity, while the results of cubic method are more noisy and slightly less intensity.

In conclusion, the propose method shows high accuracy in the phase completion task which is significantly better than the traditional methods in the industry.

5. Conclusions

This paper proposes a completion method for the incomplete phase map captured by FPP systems, based on neural networks. Firstly, we design a physical model based rendering algorithm to generate a synthetic dataset for network training. Then, we apply the advanced image completion neural network to complete the over-exposure regions based on the neighbour features. Finally, experiments are conducted to validate effectiveness of the proposed method.

Acknowledgements

This paper is supported by scientific researching fund of Software Engineering Institute of Guangzhou (ky202125).

References

[1] X. J. Jiang and D. J. Whitehouse, "Technological shifts in surface metrology," CIRP annals, vol. 61, no. 2, pp. 815–836, 2012.

[2] Y. Hu, Q. Chen, S. Feng, and C. Zuo, "Microscopic fringe projection profilometry: A review," Optics and Lasers in Engineering, p. 106192, 2020.

[3] H. Jiang, H. Zhao, and X. Li, "High dynamic range fringe acquisition: a novel 3-d scanning technique for high-reflective surfaces," Optics and Lasers in Engineering, vol. 50, no. 10, pp. 1484–1493, 2012.

[4] S. Tang and F. Gu, "Adaptive microphase measuring profilometry for three-dimensional shape reconstruction of a shiny surface," Optical Engineering, vol. 59, no. 1, p. 014104, 2020.

[5] C. J. Waddington and J. D. Kofman, "Modified sinusoidal fringe-pattern projection for variable illuminance in phase-shifting three-dimensional surface-shape metrology," Optical Engineering, vol. 53, no. 8, p. 084109, 2014.

[6] L. Zhang, Q. Chen, C. Zuo, and S. Feng, "High dynamic range 3d shape measurement based on the intensity response function of a camera," Applied optics, vol. 57, no. 6, pp. 1378–1386, 2018.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.

[8] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," Advances in neural information processing systems, vol. 25, 2012.

[9] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, "Mat: Mask-aware transformer for large hole image inpainting," arXiv preprint arXiv:2203.15270, 2022.

[10] D. Li and J. Kofman, "Adaptive fringe-pattern projection for image saturation avoidance in 3d surface-shape measurement," Optics express, vol. 22, no. 8, pp. 9887–9901, 2014.

[11] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," IEEE transactions on image processing, vol. 10, no. 8, pp. 1200–1211, 2001.

[12] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2536–2544.

[13] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5505–5514.

[14] 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), 2018, pp. 85–100.

[15] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 4471–4480.

[16] K. He, X. Chen, S. Xie, Y. Li, P. Dollar, and R. Girshick, "Masked autoencoders are scalable vision learners," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16000–16009.