Overview of Image Inpainting Techniques: A Survey

Riya Shah*, Anjali Gautam*, Satish Kumar Singh*
* *Department of Information Technology,*
Indian Institute of Information Technology Allahabad
Prayagraj, Uttar Pradesh, INDIA
riyashah2497@gmail.com, anjaligautam@iiita.ac.in, sk.singh@iiita.ac.in

***Abstract*—Images that are corrupted during transmission or improper storage, can be restored via image inpainting by filling the missing/damaged region of images such that the observer cannot perceive the inpainted region. This paper provides an overview of the traditional methods and deep learning methods which have been used for inpainting task. The traditional methods could accurately fill the missing regions when the hole size is small but fails to inpaint large sized holes and also they cannot hallucinate novel contents. With the availability of huge computation power systems, the recent advancements in deep learning methods have shown exceptional results in image inpainting tasks. However, there is still room for improvement in this task in terms of the mask size of arbitrary shape and at arbitrary locations, reducing computational resources, reducing training time, generating high quality results etc.**

***Index Terms*—Inpainting, Diffusion based techniques, Patch-based techniques, Convolution Neural Network, Partial Convolution, Generative Adversarial Network(GAN)**

I. Introduction

Reconstruction of damaged areas in photos and videos is referred to as "image inpainting". This must be accomplished in a non-detectable manner. With the invention of film and photography, the need for image inpainting increased as the photo can be deteriorated by noise during transmission or inappropriate storage. Therefore, this gave rise to digital image inpainting. Image inpainting finds its applications in many areas, like the removal of an object and filling patched region with the background pixels, photo editing, etc. Thus, algorithms that generates photo-realistic results are required.

Image inpainting has been used by researchers for a long time to provide improved results. In 2000, Bertalmio et al. [1] devised a technique based on partial differential equations, which predict pixels and fill small size holes only and have showed better results. With the advancement of technology and the increase of computing power in the age of artificial intelligence, it becomes important to develop a model using deep learning methods which can generate photo-realistic images using available large data and computation power. This paper presents the overview of different inpainting techniques and researchers working in this field can refer it. Fig. 1 shows an example of image inpainting on Paris Street View dataset.

978-1-6654-6658-5/22/$31.00 ©2022 IEEE

Fig. 1. Image inpainting on the right side of the image [2].

II. Dataset

With the increased usage of deep learning in current inpainting research; masks and data are critical components for training and evaluating the algorithms' effectiveness. In image inpainting, some most commonly used mask and image datasets are Places2, Paris Street View, CelebA, ImageNet and Oxford Building.

Places2 [3]: It is made up of images of various sceneries and categories such as airport, bedroom, streets, cathedral, etc. It comprises of almost 10 million images divided into 400 different scene groups. Each class has 5000 to 30,000 training photos in the dataset.

Paris Street View [5]: Doersch et al. created this dataset, which is based on Google Street View. A total of 10,000 photographs were collected for each city from 12 locations throughout the world, but the majority of images belong to Paris city. The resolution of each image is 936×537.

CelebA [4]: The CelebFaces Attributes Dataset (CelebA) has 202,599 celebrity facial photos with 10,177 identities, 5 landmark locations, and each with 40 binary attribute annotations cropped to a resolution of 178×218 pixels.

ImageNet [6]: It is a WordNet-based image database and contains over 100,000 synsets, with nouns accounting for the vast bulk (over 80,000). A "synset" is a significant idea in WordNet that can be defined by numerous words or phrase. To illustrate each synset, ImageNet aims to give on average 1000 images. Each image is checked for quality and annotated by a person.

Oxford Buildings [7]: This dataset contains 5062 images gathered from Flickr by searching for specific Oxford landmarks. The dataset was manually annotated to form ground

2022 IEEE Region 10 Symposium (TENSYMP) | 978-1-6654-6658-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/TENSYMP54529.2022.9864513

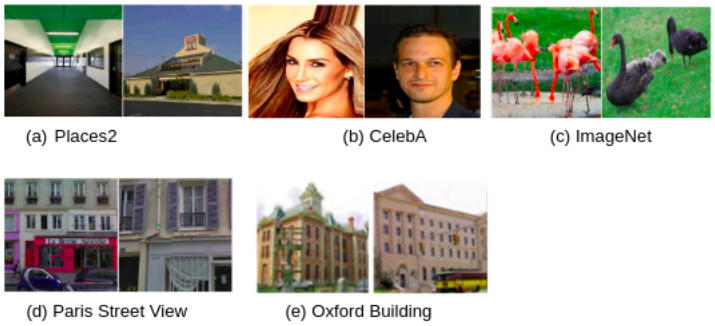

Fig. 2. Two sample images from each of (a) Places2 [3], (b) CelebA [4], (c) Paris Street View [5], (d) ImageNet [6] and (e) Oxford Building [7] datasets.

truth for eleven different landmarks. The resolution of each image is 1024×768.

Fig. 2 shows two sample images from each of the datasets described above.

III. Overview

Image inpainting fill the missing holes in the damaged image and this has been done by two different methods namely traditional and deep learning methods.

A. Traditional Methods

In this section, some traditional methods which have been used for inpainting tasks are discussed.

Bertalmio et al. [1] in 2000, put forth a technique of inpainting image that used the concept of partial differential equation. The concept of isophotes was used by the algorithm. Isophotes are curves of light with constant intensity on the surface. Information from the regions surrounding the missing regions was propagated into the missing region in the direction of isophotes.

Taking inspiration from this, Chan [8] in 2001 developed Total Variation (TV) based model. The drawback of this model is that it cannot inpaint the single object if its other disconnected parts are located far apart in the domain of inpainting. In 2001, Chan and Shen in [9] improved their model and introduced the Curvature-Driven Diffusion model (CDD) by adding isophotes' geometric information which was absent in the earlier model. The improved model was able to inpaint large holes pretty well.

In 2006, Tschumperle [10] proposed an anisrotropic smoothing algorithm using curvature-preserving Partial Differential Equations (PDEs). The thin structures of the image can be preserved with their algorithm by performing smoothing along the integral curves using Runger-Kutta interpolation. Barnes et al. [11] in 2009, proposed a nearest neighbour patch searching methodology called PatchMatch. It tries to find the best patch matches with the help of random sampling. They used Nearest-Neighbour Field (NNF) algorithm for calculating the corresponding patches.

In 2010, Xu and Sun [12] proposed a patch-based technique using Patch Sparsity. Their method was based on the patch propagation method, in which patches from the source region were propagated patch after patch into the interior of the target region. In this algorithm, the patch with the highest priority was found among the candidate patches to fill the hole. The priority was decided on the basis of Structure Sparsity. Lee et al. [13] proposed an extended version of the inpainting algorithm using segmentation of region, which uses a segmentation map. The segmented image map's boundary information to find the suitable sized candidate patch that best fits the missing region. They used a curve connection method [19] to fill the target region. In 2015, Ruzic and Pizurica [14] presented a Markov Random Field (MRF) based inpainting methodology that employs a context-aware strategy to limit the number of candidate labels per MRF node. It works by reducing each node's labels to a limited number based on the node's agreement with its labels as well as the contextual similarity between the node and the label's areas.

In 2016, Akl et al. [15] presented an algorithm inspired by the work of [20]. The algorithm consists of two steps. In the first step, it inpaints the image's structure layer. In the second step, it uses the structure layer that was obtained from step one for the inpainting process of the image. In 2017, inspired by the work of Bertalmio [1], Li et al. [16] proposed a method to inpaint large holes. They noticed that in the inpainted and

TABLE I
Traditional Methods Overview

Paper by	Year	Method Used
Bertalmio et al. [1]	2000	Partial Differentian Equation (PDE)
Chan et al. [8]	2001	Total Variation(TV)
Chan et al. [9]	2001	Curvature Driven Diffusion method (CDD)
Tschumperle [10]	2006	Anisotropic smoothing using PDEs
Barnes et al. [11]	2009	Nearest Neighbor Field (NNF)
Xu and Sun [12]	2010	Patch Sparsity method
Lee et al. [13]	2012	Region segmentation using segmentation map
Ruzic and Pizurica [14]	2015	MRF with context-aware strategy
Akl et al. [15]	2016	Two stage iterative algorithm using Markov Random Field
Li et al. [16]	2017	Partial Differential Equations, Exclusion of Abnormal Exposed Regions, Morphological Filtering
Ding et al. [17]	2018	Non Local Texture Similarity (NLTS)
Sari et al. [18]	2021	Fast Marching Method(FMM)

undamaged regions, the change in the Laplacian of the image taken along the perpendicular direction to that of the gradient is distinct. They extracted inter-channel and intra-channellocal variances by using this trait and using this trained a classifier that could recognize inpainted pixels. In 2018, Ding et al. [17] presented a technique based on texture similarity. The approach matches candidate and target patches in an image using a non-local texture similarity (NLTS) metric. The methodology chooses a patch that is centered on the contour enclosing the external border of the region to be filled as the next intended patch that is to be inpainted. It then merges them using the α- trimmed [21]to inpaint targeted patch in image. In 2021, Sari et al. [18] presented an algorithm for filling large holes by propagating structure and colour into the missing regions for inpainting large holes. Firstly, the image with large hole is segmented into small holes using structure propagation [22] step. Small holes are filled up in colour propagation step by employing Fast Marching Method FMM [23].

Table I gives the overview of these traditional methods.

B. Deep Learning Methods

Advances in deep learning, notably in Convolution Neural Networks (CNN) have been made whose one of the applications is image recognition, picture inpainting now has the missing tool. LeCun [24] introduced the Deep Learning (DL) method, which allowed inpainting of pictures using supervised image classification. The premise is that every image is associated with label. Each label is unique, and CNN learns to recognise the image-to-label mapping.

In 2016, Pathak et al. [25] introduced Context Encoders(CE) by using conditional Generative Adversarial Networks (GANs) [26]. Context encoders are encoder-decoder networks. The encoder tries to map the hidden representations and learn the relation between the pixels of missing and filled regions. From this information, the decoder tries to predict the pixel and fill in the missing region.Iizuka et al. [27] enhanced the above methodology by adding one more discriminator to assure coherency in a local image. The global discriminator focuses on the full image and ensures its coherency. The local discriminator takes into consideration only the patch of inpainted image and ensures its coherence with its surroundings by using dilated convolution layers.Li et al. [28] proposed a generative adversarial model with semantic parsing network. The generator consists of an encoder-decoder architecture [25]. The discriminator consists of two discriminators: local and global. The parsing network is primarily an encoder model [29] and aims at matching the generated image with one of the real images from the dataset. The model fails in bringing semantic coherence between the inpainted pixels and nearby pixels. In the same year, Yeh et al. [30] put forth a semantic image inpainting model using context loss and prior loss. The context loss is a weighted l1-norm difference between the image generated by the generator and the un-corrupted part of the input image. The prior loss is identical to the discriminator loss. The advantage of this methodology over the Contextual Encoders is that it does not require masking during training. However, it does not inpaint the image with holes of arbitrary sizes and locations.

In 2018, Yu et al. [31] put forth a contextual attention model. They proposed a two-stage model: first, a coarse to fine network; and second, a contextual attention model(CAM). The model has the capability to hallucinate novel content. They also introduced Spatial Discounting Loss, which is a weight matrix that consists of the value of the nearest known pixel that has an influence on the pixel to be predicted in the hole region.

In 2019, a novel approach called PEPSI (parallel extended-decoder path for semantic inpainting), a fast image inpainting methodology for semantic images using parallel decoding structure, was proposed by Sagong et al. [32]. PEPSI is made up of a single shared encoding network and a parallel decoding network with coarse and inpainting paths to reduce the amount of convolution operations. This greatly improved the traditional CAM model [27] and reduced the number of layers of convolution by almost half. In 2020, Ge et al. [33] addressed the issue of occluded face recognition and proposed an Identity-Diversity GAN (ID-GAN). The model integrates CNN face recognizer into the GAN-model. The recognizer loss, which is named as the identity-diversity loss, could collect identity-centred features which are fed back to the generator and thus improve the inpainting results.

In 2020, Mohite et al. [34], based on the work of Yu et. al. [31] proposed a method using the partial convolution layer in the coarse network before passing the output to the Contextual Attention layer. This helped not only to decrease the training time, but also improved the results obtained from the CAM model [31]. Wu et al. [35] used a network based on semantic image inpainting, which is an improvement on the model proposed in [30]. They used the Boundary Equilibrium Generative Adversarial Network (BEGAN) for the generator part of the GAN. In the discriminator, they used Self-Attention Generative Adversarial Network (SAGAN) and replaced the convolution layers with resblocks. While the model is able to achieve good results, it fails in inpainting the side-faces. Cheng et al. [36] addressed the issue of over-fitting faced by the above mentioned methods and proposed an improved GAN based method on a two-discriminator network model. The entire algorithm is made up of two parts, namely the repairing network and the discriminator network. The repairing network uses the simplified Patchmatch [11] algorithm to find the nearest similar-fit block which can be used to fill the broken part of the image. The discriminator network has global and local discriminators whose definition is the same as mentioned in [27]. In 2021, Liu et al. [37] proposed a probabilistic diverse GAN (PD-GAN) using vanilla GAN for inpainting of image. Rather than sending input images to CNN for content generation, PD-GAN commences with a noise vector, which was random, and then it was decoded. Prior information as well

TABLE II
OVERVIEW OF FEATURES, LOSS FUNCTIONS AND COMPUTATIONAL ANALYSIS OF DEEP LEARNING TECHNIQUES.

Deep Learning Techniques Overview			
Paper by	**Feature**	**Loss function used**	**GPU type and computational analysis**
Pathak et al. [25], 2016	Context Encoders (CE)	L2 loss, GAN Loss	Training time for 100,000 iterations is 14 hours
Izuka et al. [27], 2017	Contextual Encoder with Local and Global discriminator. Dilated Convolution. Poisson blending	MSE and Adversarial loss	Training time for 500,000 iterations on a single machine consisting of four K80 GPU is 2 months
Li et al. [28], 2017	GAN with Semantic Parsing Network	L2 loss, Adversarial loss.l loss , pixel-wise Softmax loss	–
Yeh et al. [30], 2017	GAN with Weighted Context Loss and Prior Loss	Context Loss ,Prior Loss	After 1.5K iterations Back-propagation was stopped
Yu et al. [31], 2018	Contextual Attention Module (CAM)	Spatially discounted reconstruction loss, WGAN-GP loss	Training time is 120 hours and model is trained on GTX-1080 Ti GPU
Sagong et al. [32], 2019	Parallel Extended Decoder Path for Semantic Inpainting (PEPSI), Region Ensemble Discriminator(RED)	L1 loss, Adversarial loss using hinge loss and spectral normalization	Model is trained on CPU Intel(R), Xeon(R) CPU E3-1245v5, GPU TITAN X
Ge et al. [33], 2020	Identity-Diversity GAN	Spatial discounted reconstruction loss, Identity-Diversity loss	–
Mohite et al. [34], 2020	Two stacked generative networks and CAM with partial convolution layers	L^1 loss, perceptual loss, style loss	AWS 64 GB RAM GPU is used and it takes 20s per image
Wu et al. [35], 2020	Generator built with BEGAN and discriminator with SAGAN	Weighted context loss , Prior loss	Intel Corei7 8700 CPU 3.2Ghz and NVIDIA GTX 1080ti GPU is used
Cheng et al. [36], 2021	GAN with patchmatch	$Loss_{local}$	Intel Corei7-7700K 4.20GHz CPU
Liu et al. [37], 2021	Probabilistic Diverse GAN (PD-GAN)	Perceptual diversity loss, SPADE, reconstruction loss, hinge loss, feature matching loss	Model is trained for 500K iterations
Suin et al. [38], 2021	Knowledge Distillation	Spatially-varying reconstruction loss, standard style loss, attention transfer loss, distillation loss	Ubuntu 16system, i7 3.40GHz CPU andan NVIDIA RTX2080Ti GPU
Sagong et al. [39], 2021	Rate-adaptive dilated convolutional layers	L1 loss, adversarial loss using hinge loss and spectral normalization	CPU Intel(R) Xeon(R) CPU E3-1245v5 and GPU TITAN X (Pascal)
Jam et al. [2], 2021	Reverse Masking Network (R-MNet)	Perceptual loss, Reverse mask loss	NVIDIA Quadro P6000 GPU and it took 7 days for training on 100 epochs
Yu et al. [40], 2022	External spatial attention (ESPA), spatially adaptive normalization (SPADE)	Perceptual Loss, Reconstruction Loss, Adversarial Loss	Model is trained on NVIDIA 2080Ti GPUs
Zeng et al. [41], 2022	Aggregated Contextual Transformation	Reconstruction loss, an Adversarial Loss, Style Loss, Perceptual Loss	–

as the region mask were injected into all decoder stages. To give prior knowledge for the generation process, the mask picture and coarse prediction were sent to SPDNorm Residual Blocks. The hard SPDNorm adjusts the probability based on the distance between the pixel and the hole boundary, whereas the soft SPDNorm learns the probability over time. Suin et al. in [38] presented a distillation-guided training methodology in which they individually guided different layers of network, so that it converges to better optima, and they also demonstrated that it can improve above mentioned methods. Their model consisted of two networks, namely, the Auxiliary Network (AN) and the Inpainting Network (IN). Masked images and ground truth images serve as input to the IN and AN networks. As guidance, supervisory signals are received by the IN network from the corresponding sub-network of AN network. This method of learning features from AN network by IN network is called Knowledge Distillation.

Sagong et al. [39] proposed a model named as Diet-PEPSI, which is an improvement over their previous work [32]. The new model drastically reduces the parameters of the network while maintaining performance by replacing dilated convolution layers with their novel rate adaptive dilated convolution layers. Jam et al. [2] have proposed a reverse masking methodology for image inpainting. One issue that persists in the above mentioned deep learning methods is the inefficiency of blending new pixels with those that are already there. Jam et al. addressed this issue in their proposed methodology, which is a combination of Wasserstein GAN and Reverse mask network (R-MNet) which is a reverse masking operator. R-MNet is a network that anticipates missing pieces of an image and keeps its authenticity by incorporating all textural and structural data. Their algorithm considered the image's overall semantic structure and predicted fine texture features that were visually plausible.

In 2022, Yu et al. [40] proposed an interactive image inpainting method that combines model priors and user assistance to inpaint the corrupted image. They proposed an External Spatial Attention Module (ESPA) to improve inpainting quality by integrating encoded and context characteristics of the corrupted image and realized spatial information with the help of a lightweight external attention method. An autoencoder built on ESPA, a semantic segmentation network, and a decoder built on spatially adaptive normalization (SPADE) [42] are the three components of the model. There are two stages of the inpainting procedure: The ESPA autoencoder generates a rough inpainting result in the first stage, which is then given to

TABLE III
PSNR AND SSIM VALUES OBTAINED USING VARIOUS DEEP LEARNING METHODS FOR DIFFERENT MASK TYPE AND SIZE

Mask Type	Mask size	Paper	Image Size	PSNR	SSIM	Dataset
Irregular	(1% - 10%)	Zeng et al. [41], 2022	512×512	34.79	0.976	Places2
	(10% - 20%)	Zeng et al. [41], 2022	512×512	29.49	0.94	Places2
	(20% - 30%)	Zeng et al. [41], 2022	512×512	26.03	0.890	Places2
	(30% - 40%)	Zeng et al. [41], 2022	512×512	23.58	0.835	Places2
	(40% - 50%)	Zeng et al. [41], 2022 Liu et al. [37], 2021	512×512 256×256	21.65 23.15	0.773 0.782	Places2
		Suin et al. [38],2021	–	25.72	0.812	Paris Street View
		Zeng et al. [41], 2022	512×512	24.06	0.834	CelebA-HQ
	(50% - 60%)	Zeng et al. [41], 2022	512×512	19.01	0.68	Places2
	˜50%	Yu et al. [31], 2018	256×256	18.91	–	Places2
	(1% - 60%)	Jam et al. [2], 2021	256×256	39.55 39.66	0.91 0.93	Paris Street View Places2
	˜80%	Yeh et al. [30], 2017	64×64	22.8	–	CelebA
	–	Wu et al. [35], 2020	64×64	22.89	0.87	CelebA
		Sagong et al. [32], 2019	256×256	24.8	0.882	Places2
		Sagong et al. [39], 2021	256×256	25.2	0.889	Places2
Regular	square (48 $\times$48)	Iizuka et al. [27], 2017 Ge et al. [33], 2020	128×128	31.0113 31.5588	0.9565 0.9598	CelebA
	center square (214×214)	Zeng et al. [41], 2022	512×512	22.1	0.84	ImageNet
	square	Pathak et al. [25], 2016	–	17.59	–	Paris Street View
		Sagong et al. [32], 2019 Sagong et al. [39], 2021	256×256	21.2 21.5	0.832 0.84	Places2
	30%	Cheng et al. [36], 2020	–	22.68	0.8685	Oxford Building
	50%	Li et al. [28], 2017	128×128	19.67	0.803	CelebA
		Yu et al. [40], 2022	256×256	28.09	0.9232	CelebA-HQ
	center	Wu et al. [35], 2020	64×64	22.89	0.87	CelebA
	–	Mohite et al. [34], 2020	512×512	40.86 40.86	– –	CelebA Places2

the semantic segmentation module to supply a semantic mask for user interaction. In stage second, A semantic decoder is used to synthesize a fine result of inpainting as directed by the customized semantic mask provided by user, ensuring that the ultimate inpainting result has the same content with the user's instructions while preserving the textures and colours restored in stage first.

Zeng et al. [41] proposed Aggregated Contextual Transformation Generative Adversarial Network. The generator consists Aggregate Contextual Transformations (AOT) block, a stack of carefully designed multiple layers for context reasoning. A discriminator is trained using a custom mask prediction task to aid in the creation of fine-grained textures. The AOT blocks, in particular, use a split-transform-merge technique. First, a block of AOT divides a standard convolution's kernel into several sub-kernels. Next, it employs each sub-kernel with varied dilation rates to the input features. Ultimately, the AOT block collects various changes from all sub-kernels. Using varied dilation rates, these three processes exploit informative faraway contexts of image. The rich patterns of interest are captured by combining results of multiple transformations which leads to better reasoning in context of the missing parts of the image.

Table II shows the main features, loss functions and the computation analysis of above discussed deep learning methods.

IV. EVALUATION METRICS

To measure the effectiveness of various developed methods for image inpainting, Peak signal-to-noise ratio (PSNR) and Structure Similarity Index Measure (SSIM) are two extensively used assessment measures.

PSNR : It is a criterion that quantifies the relationship between a signal's maximal energy and the noise that impact on its fair representation. It is a relationship between the ground truth image (original image) and the inpainted image. The image quality is directly related to PSNR, therefore, higher the value, the better the quality of the image.

SSIM : It measures the similarity between two images, and values fall in the range of negative 1 and positive 1. A value of positive 1 indicates that the given two images are similar, while a value of negative 1 suggests that the given two images are very different. These numerals are often adjusted to fall between 0 and 1, inclusive.

Table III indicates PSNR and SSIM values achieved on different datasets by the above mentioned deep learning methods.

V. CONCLUSION

This paper gives an overview and recognises some of the prominent image inpainting methods. We have divided methods of image inpainting into Traditional and Deep Learning methods. Traditional methods are efficient when the size of the holes is small. They are unable to hallucinate novel contents that are absent in the original image. The results produced by them are blurry and not consistent with their neighbors. Deep Learning methods can solve the above mentioned problems of traditional methods. They are able to generate images that are semantically coherent and can hallucinate novel contents. But they require a lot of external information and are to be trained on multiple datasets. The training time requirement is also enormous. Much effort has been expended into reducing the model's training time. Some researchers have also worked

to address the problem of inpainting with masks of arbitrary size and shape, and there is scope for improvement in the form of improving the clarity of results, reducing training time, etc.

REFERENCES

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

[2] J. Jam, C. Kendrick, V. Drouard, K. Walker, G.-S. Hsu, and M. H. Yap, "R-mnet: A perceptual adversarial network for image inpainting," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2714–2723.

[3] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[4] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[5] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. Efros, "What makes paris look like paris?" *ACM Transactions on Graphics*, vol. 31, no. 4, 2012.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[7] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[8] T. Chan, "Local inpainting models and tv inpainting," *SIAM Journal on Applied Mathematics*, vol. 62, no. 3, pp. 1019–1043, 2001.

[9] T. F. Chan and J. Shen, "Nontexture inpainting by curvature-driven diffusions," *Journal of Visual Communication and Image Representation*, vol. 12, no. 4, pp. 436–449, 2001.

[10] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving pde's," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 65–82, 2006.

[11] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, p. 24, 2009.

[12] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1153–1165, 2010.

[13] J. Lee, D.-K. Lee, and R.-H. Park, "Robust exemplar-based inpainting algorithm using region segmentation," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 553–561, 2012.

[14] T. Ružić and A. Pižurica, "Context-aware patch-based image inpainting using markov random field modeling," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 444–456, 2014.

[15] A. Akl, E. Saad, and C. Yaacoub, "Structure-based image inpainting," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2016, pp. 1–6.

[16] H. Li, W. Luo, and J. Huang, "Localization of diffusion-based inpainting in digital images," *IEEE transactions on information forensics and security*, vol. 12, no. 12, pp. 3050–3064, 2017.

[17] D. Ding, S. Ram, and J. J. Rodríguez, "Image inpainting using nonlocal texture matching and nonlinear filtering," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1705–1719, 2019.

[18] I. N. Sari, E. Horikawa, and W. Du, "Interactive image inpainting of large-scale missing region," *IEEE Access*, vol. 9, pp. 56 430–56 442, 2021.

[19] J. C. Hung, C.-H. Huang, Y.-C. Liao, N. C. Tang, and T.-J. Chen, "Exemplar-based image inpainting base on structure construction," *Journal of Software*, vol. 3, no. 8, pp. 57–64, 2008.

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

[21] J. Bednar and T. Watt, "Alpha-trimmed means and their relationship to median filters," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 1, pp. 145–153, 1984.

[22] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum, "Image completion with structure propagation," in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 861–868.

[23] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6721–6729.

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

[25] 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.

[26] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[27] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.

[28] Y. Li, S. Liu, J. Yang, and M.-H. Yang, "Generative face completion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3911–3919.

[29] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang, "Object contour detection with a fully convolutional encoder-decoder network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 193–202.

[30] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5485–5493.

[31] 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.

[32] M.-c. Sagong, Y.-g. Shin, S.-w. Kim, S. Park, and S.-j. Ko, "Pepsi: Fast image inpainting with parallel decoding network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 360–11 368.

[33] S. Ge, C. Li, S. Zhao, and D. Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 10, pp. 3387–3397, 2020.

[34] T. A. Mohite and G. S. Phadke, "Image inpainting with contextual attention and partial convolution," in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2020, pp. 1–6.

[35] C. Wu, Y. Xian, J. Bai, and Y. Jing, "Semantic image inpainting based on generative adversarial networks," in *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*. IEEE, 2020, pp. 276–280.

[36] Y. Chen, H. Zhang, L. Liu, X. Chen, Q. Zhang, K. Yang, R. Xia, and J. Xie, "Research on image inpainting algorithm of improved gan based on two-discriminations networks," *Applied Intelligence*, vol. 51, no. 6, pp. 3460–3474, 2021.

[37] H. Liu, Z. Wan, W. Huang, Y. Song, X. Han, and J. Liao, "Pd-gan: Probabilistic diverse gan for image inpainting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9371–9381.

[38] M. Suin, K. Purohit, and A. Rajagopalan, "Distillation-guided image inpainting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2481–2490.

[39] Y.-G. Shin, M.-C. Sagong, Y.-J. Yeo, S.-W. Kim, and S.-J. Ko, "Pepsi++: Fast and lightweight network for image inpainting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 252–265, 2021.

[40] W. Yu, J. Du, R. Liu, Y. Li *et al.*, "Interactive image inpainting using semantic guidance," *arXiv preprint arXiv:2201.10753*, 2022.

[41] 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.

[42] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2337–2346.