Image inpainting using Deep learning methods

Aiswarya Annem
Department of Computing Sciences
Texas A&M University-Corpus Christi
Corpus Christi, Texas,US
aannem1@islander.tamucc.edu

Dr. Longzhuang Li
Department of Computing Sciences
Texas A&M University-Corpus Christi)
Corpus Christi, Texas,US
longzhuang.li@tamucc.edu

***Abstract*—Image inpainting is the process of rebuilding the missing parts of the image, so the viewer cannot judge that these areas have been restored. This technique is used to remove unwanted objects from an image or restore damaged parts of old photos and even image denoising can be done using inpainting. In this paper,we propose a generative adversarial network based method for object removal and scene generation. In future work this idea can be useful for computer vision and pattern recognition tasks such as image denoising and object detection.**

***Index Terms*—Image inpainting, object removal, image regeneration, scene generation**

I. Introduction

Image inpainting [4] refers to the process of filling in missing data in a designated region of the visual input. The object of the process is to reconstruct missing parts or damaged images in such a way that the inpainted region cannot be detected by a casual observer. Applications range from the reconstruction of missing blocks introduced by packet loss during wireless transmission, reversing of impairments, such as cracks, scratches, and dirt, in scanned photographs and digitized artwork images, to removal/introduction of image objects such as logos, stamped dates, text, persons, and special effects on the scene. Typically, after the user selects the region to be restored, the inpainting algorithm automatically repairs the damaged area by means of image interpolation [14].

In other words, Image Inpainting is the process of filling missing regions of an image and it is a crucial stage in lot of images editing task. It is used to fill the missing regions after eliminating unnecessary elements from an image. For example,humans have an incredible ability to spot irregularities in the visual world. As a result,the filled portions must be believable from a perceptual standpoint. Among other things, the lack of fine structure in the filled region is a giveaway that something is amiss, especially when the rest of the image contain sharp details [9]. Image inpainting divided into two stages:edge generation [11] and image completeion. Edge generation in the missing regions is entirely focussed on hallucinating edges. Image completion network calculates the RGB pixel intensities of the missing regions based on hallucinated edges and Visually, the RGB pixel intensities are uniform. Both networks use deep feature-based losses to ensure perceptually accurate outcomes

Deep learning algorithms have recently achieved surprising success when it comes to image inpainting. These approaches use learned data distribution to fill in the missing pixels. They can construct coherent structures in the missing regions, which were previously thought to be unachievable using standard methods. While these algorithms can build missing regions with meaningful structures, the reconstructed regions are frequently hazy or contain artifacts, implying that algorithms struggle to effectively reconstruct high-frequency information.

Image inpainting has been addressed for over two decades, and there are two types of approaches: Machine learning and Deep learning-based methods. In this edge-to-image method: To create an edgeness map of an incomplete image, we use a deep network-based edge detector, then fill in the blank areas in the edgeness map, and finally construct the missing pixels with the help of the complete edgeness map. Especially, This method Shunxin Xu, et al. [20] is able to reproduce image structures that appear natural and the plausible and proposed method fails in case the missing area is very large. Compared with other methods can better remove the artifacts at the boundary, mainly due to the Bi-directional Skip Connections mechanism in the dense multi-scale fusion of local binary patterns [10] (DMFLBP)learning network guidance and repair network but the main semantic information of the missing area is repaired, but it will be incoherent at the defect boundary. This method inhua Liu et al. [10] has a good repair effect, but some fine textures at the boundary are very blurry and produce artifacts. this produces better repair results through edge contours, but some broken or blurred edges can also be observed, and artifacts are more obvious at the border. The principal idea of the proposed Dynamic selection network(DSNet) Ning Wang et al. [19]is to distinguish the corrupted region from the valid ones throughout the entire network architecture, which may help make full use of the information in the known area. Despite the fact that the suggested DSNet is effective at image inpainting, it still has flaws. To begin with, the suggested method creates new convolution and normalization processes without the need for optimization. As a result, with careful optimization, the processing time can be lowered. Second, for heavily damaged regions, the proposed technique is unable to restore the original information. Image editing activities such as face editing may benefit from the proposed solution. However, it is unable to assist with specialized jobs that need the recovery of the correct image, such as suspicious face restoration.

In this research, we are proposing a method which will try to overcome the limitations presented above. Most of the state of art methods make use of CelebA and Place2 dataset. So, in this research we will evaluate the proposed method on these state-of-art datasets for image inpainting task. The main contribution our paper will be a technique proposed by Kaiming He in ResNet [6] is used to to strengthen the predictive ability of the generator and to prevent the gradient vanishing caused by the deep network and it will be more realistic. Instead of using the original cross-entropy loss, we employ the Wasserstein GAN [21]loss to overcome the challenge of training adversarial networks and increase the accuracy of completed regions. Most important use cases of this proposed method are object removal and scene generation. Also, it can be used in forensics to regenerate image based evidences.

This paper is organized in 4 different sections mainly: (i) Related Works (ii) Proposed Method (iii) Experiments and Results (iv) Conclusion.

II. Related Works

This section discusses about the existing methods and research going on in the field of image inpainting it shows how much in depth the research has been concluded as of now and what are the advantages and disadvantages of those methods:

In this paper X.Zhou et. al [22] proposes a segmentation reconstruction network and image inpainting network to address global structure prediction and texture synthesis. In image inpainting a graph-based network is used to represent relationships existed in a corrupted image containing an intra relationship for pixels in the same semantic region and interrelationship in different semantic a, as well as contrastive loss is also proposed [12] to make relation network training easier. After Quantitative comparisons, it was able to achieve PSNR values of 28.61 and 24.16 on CelebA-HQ and DeepFashion datasets respectively and it outperforms when compared with Edgeconnect, GatedConv, StructureFlow. Many deep learning models have been proposed to handle image inpainting problems by learning vast amounts of data, and many techniques employ monotonous normalization and static convolution, which consider pixels with fixed grids. however, these are ineffective due to random image corruption. In this paper [19], Ning Wang et al. proposes a dynamic selection network (DSNet) that includes both VMC (validness migratable convolution) and RCN (regional composite normalization) modules to solve convolution stage problems and RCN (regional composite normalization) modules to solve normalization stage problems. Even though the proposed DSNet is effective in inpainting images, it has flaws.

The main benefit of generative adversarial network (GAN) [2] [13] is that it restores semantic images [7] such as human face images that have problems with gradient vanishing and model collapse during training. To overcome these limitations Chong Han et al. [5] proposed GAN(generative adversarial network) with evolutionary generators and applied it to face image inpainting and to keep the model training process stable, EG-GAN evolves the generator network and uses two mutation functions as a training target, to modify the parameter of the generator network and to generate offspring generators via crossover with the help of the matcher the discriminator to criticize the image that was generated. Experiments on the CelebA-HQ and CelebA face picture datasets reveal that EG-GAN successfully solves the gradient vanishing problem, performs stable and efficient training, and creates visually reasonable images.

In this paper [8] , the Jireh Jam et al. proposed Reverse Masking Network(R-MNet) [23]with Wasserstein GAN to overcome some issues while blending the missing pixels with visible ones. with this model image details will be preserved on high-resolution images. Here CelebA-HQ test data is used and with the help of reverse matrix operands of the mask is advantageous to image inpainting. After quantitative comparison, it was able to achieve PSNR value of 40.40 on CelebA-HQ dataset. Because other models are biased towards the trained Images, many Deep learning-based image inpainting methods perform better except for untrained images. To overcome this disadvantage, the author offers AdaFill, a simple image inpainting technique with test-time adaptation. Chajin Shin et al. [16] goal is to fill in the hole region more naturally than pre-trained inpainting models given a single out-of-distributed test image. Because natural images have considerable internal similarities, we treat the remaining valid parts of the test image as additional training cues to attain this goal. Proposed network can directly leverage externally learned image priors from the pre-trained features as well as the internal priors of the test image as a result of this test-time adaptation. AdaFill beats other models on various out-of-distribution test photos, according to the results, and ZeroFill which is not pre-trained also outperforms all other models in some cases.

In this paper Yong-Goo Shin et al. [17], researchers propose PEPSI, a novel image inpainting model that uses a collaborative learning method to overcome the limitations of the two-stage coarse-to-fine network. We compared qualitative and quantitative data. on the data sets CelebA-HQ and Place2. PEPSI not only outperforms traditional techniques in terms of performance, but it also greatly decreases computing time via a parallel decoding path [3] and an excellent joint learning scheme, according to the results. Additionally, they proposed Diet-PEPSI, which aggregates global contextual information with minimal hardware costs using unique rate-adaptive convolutional layers. Diet-PEPSI keeps PEPSI's performance while reducing hardware costs dramatically, making hardware implementation easier. Both networks are trained using the recommended RED and produce visually credible responses in both square and irregularly shaped holes. As a result, the proposed methods are expected to be widely used in a variety of applications, including image generation, style transfer, and image editing. Shunxin Xu et al. [20] [18] proposed a CNN-based method for image inpainting in two steps using an end-to-end generative network [1]in this paper. The first step is to return the edgeness map to its original state the image's missing portion, and the second step is to complete it with the help of the [12]edge inpainting outcome, inpainting work.

In this approach, our suggested solution restores both image structure and texture effectively. Repairing the edgeness map, on the one hand, could aid our edge-based image inpainting network in generating more believable structures. Textures, on the other hand, can be created with the help of the recovered edgeness map, which reduces mismatch mistake. As a result, we produce superior outcomes than state-of-the-art approaches, as the extensive experimental data reveal.

III. Proposed Work

This Section First discusses the Overview of the framework and some basic symbols, and then the objective function. Finally, we propose the algorithm for our training model.

A. Overview

The following is how the task of picture inpainting is formalized: Partially mask an image x to get an incomplete input image x_m, with r denoting the actual region. To create an image f, we discover a function $f(x_m)$. Image inpainting's goal is to get the created image $f(x_m)$ as near to the original complete image x as possible. The following is an example of formal representation:

$$f = arg\min_{f} ||f(x_m) - x||_2^2$$

The architectural framework for image inpainting is described in following paragraphs were we explain on what each component is used for as shown in Figure 1. Our model's goal is to fill in the missing regions of an incomplete image so that the image can be completed. Visually and conceptually, the entire image is plausible.

B. Generator

Inpainting is one of many issues with image production. We employ an auto-encoder as the model's generator to tackle this problem. There are two networks in the auto-encoder: an encoder and a decoder. In this research, we use an encoder to encode the image to be repaired into code, then use a decoder to reconstruct and generate the corrected image. In contrast to the traditional AutoEncoder architecture, we introduce skip-connection between the respective layers of the encoder and decoder sections to prevent the network layer from deteriorating owing to network layer deepening. Skip-connection [18] can ensure that the decoding stage of the equivalent resolution can use the output of the low-level coding stage to supply the decoder with part of the structural feature information lost during the encoder down sampling phase, hence improving the generator's structure prediction capabilities. The encoder in this paper employs a multi-layer convolution layer architecture. The decoder's design is symmetrical to the encoder's, with transposed convolution layers.Instead of fully connected layers, we use four layers of dilated convolution between the encoder and the decoder.

C. Discriminator

The generator is in responsible of inpainting the image's missing or masked parts. The generator, on the other hand, cannot guarantee that the regions generated are correct or consistent with the original image. This work employs the discriminator as a binary classifier to determine whether the image comes from genuine data distribution or is generated by the generator, in order to ensure that the generated image is considerably more realistic. The discriminator also aids the generator in producing more realistic visuals in order to deceive the discriminator.

The discriminative network employs two CNN architectures as local and global discriminators. To begin, the local discriminator determines if the missing part's result is semantically correct. If the missing portion is the nose, for example, the local discriminator must determine if the completed part is indeed the nose. The missing or obscured part of the original image, as well as the part generated by the generator, are both inputs to the local discriminator. We use channel splicing to input them into the local discriminator. Because we're using the CNN architecture, the local discriminator's output is a scalar, indicating whether the produced area is true (based on real data distribution) or false (generated by the generator)However, relying solely on the local discriminator is insufficient.

Despite the fact that the output is only partially correct, the overall coherence is deemed to be successful [1]. To determine the degree of coherence between the generated region and the original image, we employ the global discriminator. The global discriminator's inputs are split into two categories, similar to the local discriminator's: the original image as ground truth and the whole restored image generated by the generator. Similarly, the channel stitches two images into the global discriminator, and the output is a scalar that indicates the global discriminator's level of confidence for the complete image after completion, as well as whether it is semantically coherent.

D. Loss Function

We'll look at each module's loss function [18] because this paper only utilizes one generator and two discriminators. We train the generator in this paper by minimizing the reconstruction loss L_r In this paper, the Lr norm loss function is chosen over the L_1norm loss function. Because the L_2 norm penalizes outliers, it is appropriate for inpainting tasks; nonetheless, the L_2 norm's robustness is insufficient. The reconstruction loss is defined as formula

$$L_r = ||G(x_m) - x||_2^2$$

Adversarial loss is required for GAN training and is becoming more popular in many creative jobs; less adversarial loss indicates that the generator has more power to fill the holes. We employ the Wasserstein GAN [21] loss and global and local discriminators for stable training. The Wasserstein GAN loss is defined as

$$\min_{G}\max_{D} V(D, G) = E_{X\sim Pdata(x)}[D(x)] - E_{z\sim p_z(z)}[D(G(z))]$$

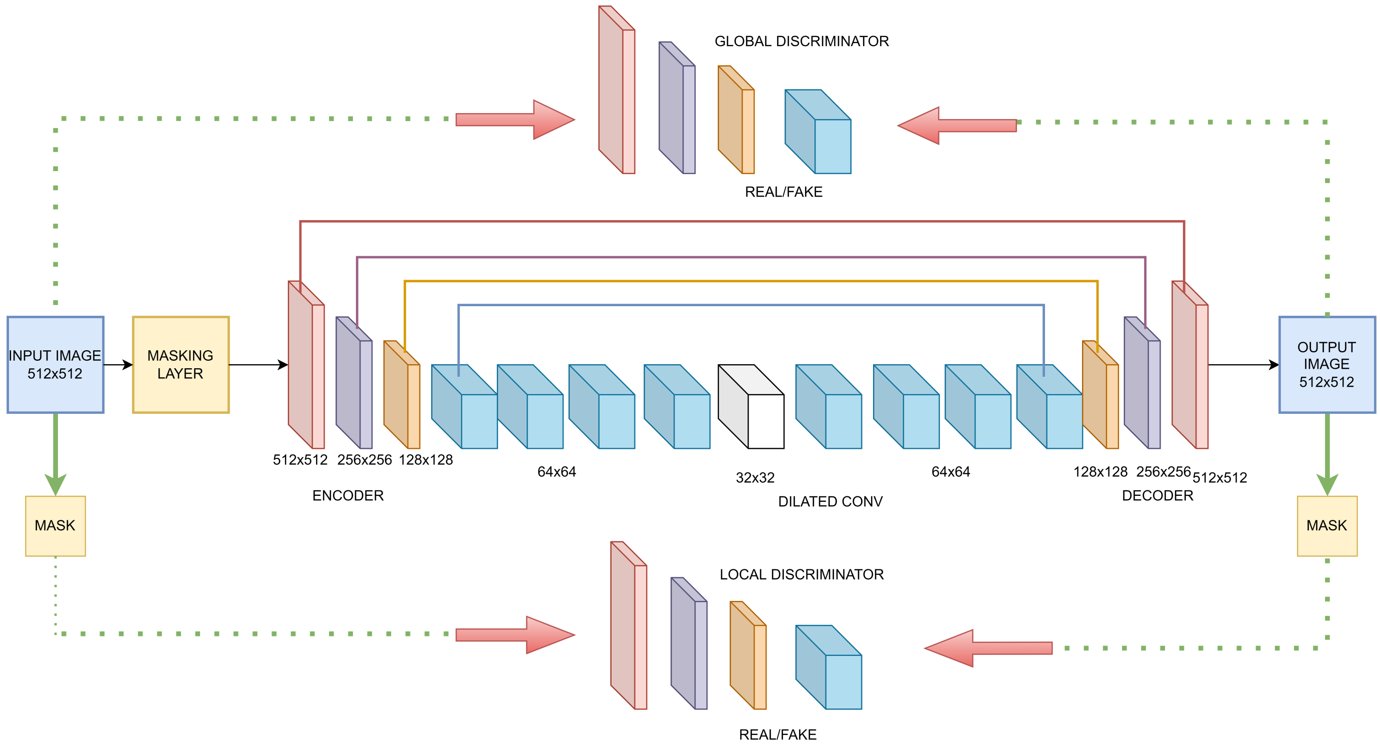


Fig. 1: Architecture of Proposed Model

E. Training

In this paragraph, we present an image inpainting algorithm for training. In each iteration, the mini-batch training approach is utilized to obliterate the dataset image. To begin, we choose a small sample of x pictures from the training data and mask them with random holes. Then we get a mini-batch of masked images z, real areas before they're masked r, and masks m, with $z = x \bigodot m$,where $\bigodot$ representing element-wise multiplication.

The generator [5] is then trained s times with l_r loss. After training the generator, we repair it and use L_{global} and L_{local} to train discriminators t times. Finally, we use joint loss L to train the joint model. Input z into the model, and the expected images are output as c. We generate the final inpainting pictures $x_i = z + c \bigodot (1 - m)$by combining the masked regions of c and z.

To evaluate this model, we have chosen Peak signal to noise ratio(PSNR) [15] as the standard by which most similar methods are evaluated against. PSNR is the ratio between the maximum possible power of an image and the power of corrupting noise that affects the quality of its representation. To estimate the PSNR of an image, it is necessary to compare that image to an ideal clean image with the maximum possible power. The PSNR is defined as:

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

Another criterion we have decided to measure is Structural similarity index(SSIM) [15]. SSIM is another measurement tool used to measure the quality of imperceptibility in steganographic images and SSIM was built based on three main factors, namely luminance, contrast, and structure. The SSIM is define as:

$$\text{SSIM} = l(x,y) * c(x,y) * s(x,y) \tag{1}$$

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2a}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2b}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{2c}$$

F. Dataset

To learn and assess our model, we use celebA, CelebA-HQ and Places2 Datasets.There are 202599 RGB color facial images in the CelebA dataset.For training, we use 100k images and for testing, we use 1000 images. Places2, a dataset that contains over 8,000,000 images from over 365 scenes collected from the natural world. For Places2 we use origin splits for validation and testing.CelebA-HQ is a high-definition human face dataset created by Nvidia in 2018. It contains 24102 photos for training and 2942 images for testing, all of which are 512 × 512 pixels in size. We map Celeb-19 HQ's fine segmentation labels to 11 categories, including face, eyes, lip, eyeglasses, and so on. Apart from the inpainting results, we use the PSNR and SSIM indices to estimate our model in our experiments.

G. Difference from existing methods

We are proposing a method that uses Generative Adversarial Networks(GAN) which mainly compresses a generator and a discriminator. several methods use convolution neural networks for the use case of image inpainting but Generative Adversarial Networks are not very well explored to address the problem of image inpainting. so we will be exploring GAN in image inpainting. Our proposed method will include one generator and two discriminators. In proposed method the generator in our case will be responsible for filling out the missing regions. We use the two discriminators which are global discriminator which will be determining whether the repair outcome is consistent throughout the process, while the local discriminator will be determining whether the repaired area is correct. The generator architecture is an autoencoder that attempts to match the picture input to the encoder with the image output from the decoder as closely as possible. The difference from our method to normal use cases is instead of using original GAN, which is very difficult to train and the loss of the generator and discriminator may not be indicative of the training process. which is very inconsistent in terms of quantifying the result. Further to address this issue we are using Wasserstein Generative Adversarial Networks(WGAN) [21] which is introduced in this paper, it is not explored in image inpainting and is very rarely used and it is a new iteration of GAN which works very well for stable training and also it can significantly reduce processing time with great optimization of the loss function. Thirdly we are using the technique proposed by Kaiming He in ResNet [6] to improve the predictive ability of the generator and improving the quality of repaired photos and solve gradient vanishing. The advantages this approach provides a realistic and semantically coherent images with high-resolution facial image inpainting which preserve color and maintain realism on a reconstructed image To summarize, no solution exists that can inpaint all sorts of image distortion, many learning techniques yield promising outcomes but they are also not perfect. So what we proposing can also have defects of its own but We hope this work paves the way for further development of inpainting algorithms.

H. Performance of existing Methods

Here we compared the performance of several existing methods which are Generative Adversarial Network with Evolutionary Generators(EG-GAN),Contrastive Relation Network(CRN), Edge-Aware Dual Branch Generative Adversarial Network (EDBGAN),Reverse Masking Network(R-Mnet) [8],semantic-aware context aggregation module(SACA), Dense Multi-Scale Fusion of Local Binary Patterns (DMFLBPN) and Learnable Edge-Attention Maps(LEAM) using PSNR and SSIM metrics for the Places2 Dataset and Celeb A dataset as described in Table 1 above.

TABLE I: Performance Comparison

Method \ Dataset	Metrics	Places2	Celeb A	CelebA-HQ
EG-GAN [5]	PSNR	-	-	34.16
	SSIM	-	-	0.94
CRN [22]	PSNR	-	-	28.61
	SSIM	-	-	0.901
EDBGAN [2]	PSNR	-	29.156	-
	SSIM	-	0.955	-
LEAM [18]	PSNR	22.26	-	-
	SSIM	0.712	-	-
DMFLBPN [10]	PSNR	-	-	28.07
	SSIM	-	-	0.893
R-Mnet [8]	PSNR	39.66	-	40.4
	SSIM	0.93	-	0.94
SACA [7]	PSNR	-	25.2	-
	SSIM	-	0.85	-

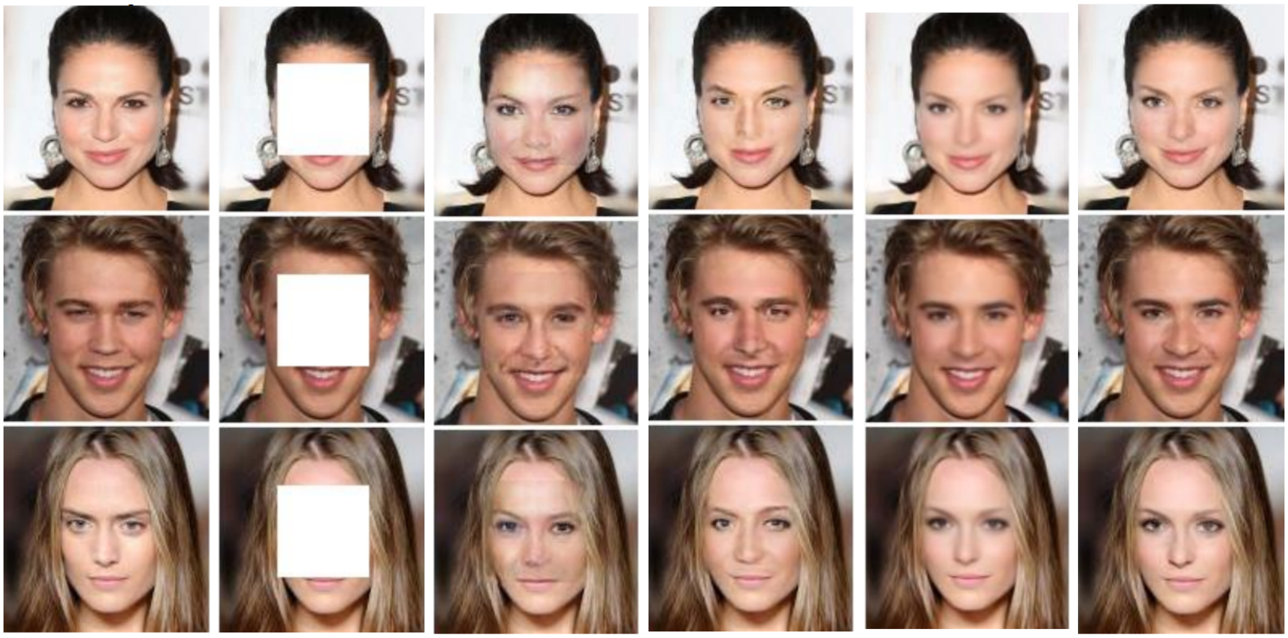
IV. Experiments

CelebA-HQ is the dataset used in this experiment. Nvidia released CelebA-HQ in 2018, a high-definition human face dataset that allows photographs in CelebA to be scaled up to 1024*1024. Our experiment used a face image with a pixel size of 512x512, a training set of 24102 images, and a test set of 2942 images. Two GTX 1080 Ti GPUs are active at the same time, completing the 60-generation cycle.

In our experiments, we employ the PSNR and SSIM indices to estimate our model in addition to the inpainting findings. Considering the importance of visual and semantic coherence; we used our test dataset to conduct a qualitative comparison. To begin, we used a WGAN technique. On the images, we saw an induced pattern and a pitiful color. To manage the induced pattern and match the luminance of the original images, we used dilated convolution, Kaiming He technique [6], and end-to-end training with the Wasserstein-perceptual loss function. To analyze the performance objectively, we chose some popular image quality metrics such as Peak Signal to Noise Ratio (PSNR) and SSIM.

The outcomes of our experiment compared to the state of the art for image inpainting using our model are shown in the table. The greater the number for PSNR and SSIM, the closer the image quality is to ground-truth. In comparison to other state-of-the-art methodologies, our model produces comparable PSNR and SSIM values, which can be improved further with further iterations, implying that our model provides more accurate predictions than the state-of-the-art inpainting methodology.

The state-of-the-art is outperformed by our suggested WGAN with dilated convolution and Kaming He method, which is trained end-to-end with Wasserstein-perceptual loss function. Our algorithm can forecast missing pixels of the binary mask regions on the image by learning the end-to-end mapping of input images from a large-scale dataset. Our model learns and recognizes missing pixels in the input and encodes them as feature representations, which can then be reconstituted by the decoder. In ResNet, the Kaiming He [6] approach aids in the forward transmission of picture features and the backward propagation of local minimums. For Image inpainting, our investigations suggest that the Kaiming



(a) Real (b) Input (c) CRN (d) EG-GAN (e) R-Mnet (f) OURS

Fig. 2: Qualitative comparisons on CelebA-HQ dataset with existing methods

He [6] Technique paired with Wasserstein-perceptual loss is beneficial. In Figure, we have visually compared our suggested technique to the state of the art. Experiments with normal convolutions were used to test the effectiveness of our network. We discovered, as illustrated in Figure, that the images created included checkboard artifacts with pitiful visual similarity to the source image. We combined the Kaiming He [6] technique with dilated convolution and our novel loss function to produce better outcomes than were semantically reasonable and realistic in every way.

We used the WGAN to investigate the impact of varying mask sizes on inpainting results. We used eight different square masks. As the mask size grows larger, SSIM and PSNR fall to some amount. SSIM, on the other hand, does not show the same declining trend as PSNR. As a result, even if W-GAN can't reach high pixel-level performance for a larger mask, it can achieve higher structural restoration. We also conducted an experiment with a face that wore spectacles, as indicated. Meanwhile, we choose some images from different datasets to test our model's generalization capabilities. All of these findings suggest that our model is capable of generating images that are unique from the training set, which is critical for avoiding model collapse. Finally, as indicated in the table, we compare our model to other image inpainting models.

TABLE II: Performance Comparison with existing methods

Method \ Dataset	Metrics	CelebA-HQ
CRN [22]	PSNR	28.61
	SSIM	0.901
EG-GAN [5]	PSNR	34.16
	SSIM	0.92
R-Mnet [23]	PSNR	40.4
	SSIM	0.94
Our Method	**PSNR**	**35.6**
	SSIM	**0.92**

V. Conclusion and Future Work

In this paper, Our model can generate images that are semantically and visually plausible while maintaining facial realism. This was achieved by employing a network topology that allows each block's receptive field to be enlarged to capture more data and pass it to the appropriate deconvolutional blocks. The proposed model has proved the effectiveness of gathering and refining contextual information when WGAN is combined with the Kaiming He [6] approach. PSNR, and SSIM, the results reveal that our suggested model produced comparable and more realistic outcomes when compared to EG-GAN [5], CRN [22], and DMFLBPN [10]. We plan to extend our model in the future to deal with the task of image inpainting with complex structural information missing, as well as compare our model to more state-of-the-art solutions. object removal and denoising can be accomplished with the proposed inpainting approach.

References

[1] Cong Chen, Amos Abbott, and Daniel Stilwell. Multi-level generative chaotic recurrent network for image inpainting. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3625–3634, 2021.

[2] Minyu Chen and Zhi Liu. Edbgan: Image inpainting via an edge-aware dual branch generative adversarial network. *IEEE Signal Processing Letters*, 28:842–846, 2021.

[3] Ye Deng and Jinjun Wang. Image inpainting using parallel network. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1088–1092, 2020.

[4] Borko Furht, editor. *Image Inpainting*, pages 315–316. Springer US, Boston, MA, 2006.

[5] Chong Han and Junli Wang. Face image inpainting with evolutionary generators. *IEEE Signal Processing Letters*, 28:190–193, 2021.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[7] Zhilin Huang, Chujun Qin, Ruixin Liu, Zhenyu Weng, and Yuesheng Zhu. Semantic-aware context aggregation for image inpainting. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2465–2469, 2021.

[8] Jireh Jam, Connah Kendrick, Vincent Drouard, Kevin Walker, Gee-Sern Hsu, and Moi Hoon Yap. R-mnet: A perceptual adversarial network for image inpainting. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2713–2722, 2021.

[9] Siyuan Li, Lu Lu, Zhiqiang Zhang, Xin Cheng, Kepeng Xu, Wenxin Yu, Gang He, Jinjia Zhou, and Zhuo Yang. Interactive separation network for image inpainting. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1008–1012, 2020.

[10] Xinhua Liu, Lihua Gan, Xiaolin Ma, and Hailan Kuang. Face image inpainting network based on dense multi-scale fusion of local binary patterns. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, volume 5, pages 2581–2585, 2021.

[11] Bojie Ma, Xiaowei An, and Nongliang Sun. Face image inpainting algorithm via progressive generation network. In *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pages 175–179, 2020.

[12] Xin Ma, Xiaoqiang Zhou, Huaibo Huang, Zhenhua Chai, Xiaolin Wei, and Ran He. Free-form image inpainting via contrastive attention network. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9242–9249, 2021.

[13] Utkarsh Maheshwari, Venkata Pavan Kumar Turlapati, and Usha Kiruthika. Lucid-gan: An adversarial network for enhanced image inpainting. In *2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 1–6, 2021.

[14] Zahra Nabizadeh, Ghazale Ghorbanzade, Nader Karimi, and Shadrokh Samavi. Image inpainting by adaptive fusion of variable spline interpolations. In *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, pages 1–5, 2020.

[15] Muhammad Ali Qureshi, Mohamed Deriche, Azeddine Beghdadi, and Asjad Amin. A critical survey of state-of-the-art image inpainting quality assessment metrics. *Journal of Visual Communication and Image Representation*, 49:177–191, 2017.

[16] Chajin Shin, Taeoh Kim, Sangjin Lee, and Sangyoun Leey. Test-time adaptation for out-of-distributed image inpainting. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2009–2013, 2021.

[17] Yong-Goo Shin, Min-Cheol Sagong, Yoon-Jae Yeo, Seung-Wook Kim, and Sung-Jea Ko. Pepsi++: Fast and lightweight network for image inpainting. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):252–265, 2021.

[18] Liujie Sun, Qinghan Zhang, Wenju Wang, and Mingxi Zhang. Image inpainting with learnable edge-attention maps. *IEEE Access*, 9:3816–3827, 2021.

[19] Ning Wang, Yipeng Zhang, and Lefei Zhang. Dynamic selection network for image inpainting. *IEEE Transactions on Image Processing*, 30:1784–1798, 2021.

[20] Shunxin Xu, Dong Liu, and Zhiwei Xiong. E2i: Generative inpainting from edge to image. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1308–1322, 2021.

[21] Yue Zhao, Lijun Liu, Han Liu, Guo Xie, and Youmin Zhang. Face inpainting based on improved wgan-modified. In *2020 International Symposium on Autonomous Systems (ISAS)*, pages 7–11, 2020.

[22] Xiaoqiang Zhou, Junjie Li, Zilei Wang, Ran He, and Tieniu Tan. Image inpainting with contrastive relation network. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 4420–4427, 2021.

[23] Manyu Zhu, Dongliang He, Xin Li, Chao Li, Fu Li, Xiao Liu, Errui Ding, and Zhaoxiang Zhang. Image inpainting by end-to-end cascaded refinement with mask awareness. *IEEE Transactions on Image Processing*, 30:4855–4866, 2021.