Image Reconstruction using Near and Far Receptive Fields

G. Rajasekaran[1] · J. Madhushree[1] · S. Mahalakshmi[1] · J.S. Mahasri[1]

Abstract

Image Inpainting improved a lot with the help of algorithms in deep learning. Image Inpainting will be very useful if the missing regions present in the document or an important image. Most widely used networks include an encoder-decoder architecture (sometimes with skip connections, allowing layers to skip layers) and a sizeable receptive field, or one that is greater than the picture resolution. For picture inpainting jobs, the size of the surrounding areas needed to fill in different sorts of missing regions vary, and a very large receptive field is not always the ideal choice, especially for the neighbouring buildings and textures. Inpainting will be hindered by the broad receptive field's tendency to create more unsatisfactory completion outcomes. We developed a ground-breaking three-stage inpainting framework with partial, small and large refinements based on these insights, which rethinks the picture inpainting process from the unique viewpoint of the receptive field. We first employ an encoder-decoder network to gather rough early results. Then, to carry out the small refinement, we introduce a shallow deep model with a small receptive field, which can also lessen the impact of distant undesirable completion outcomes. To carry out the large refinement, we finally propose an attention-based encoder-decoder network with a large receptive field. According to test results, our method performs better than the state of currently available technology for image inpainting on two popular datasets that are open to the public.

Keywords Image inpainting · Deep learning · Receptive field

✉ G.Rajasekaran

rajasekaran@mepcoeng.ac.in

J. Madhushree

madhushreejeyaram_it@mepcoeng.ac.in

S.Mahalakshmi

mahamahimasaravanan_it@mepcoeng.ac.in

J.S. Mahasri

js.mahasri01_it@mepcoeng.ac.in

[1] Department of Information Technology, Mepco Schlenk Engineering College, Sivakasi,Tamil Nadu, India

1 Introduction

Image reconstruction is also called as Image Inpainting. The reconstructed image needs to be realistic, and in order to efficient in-paint, the appropriate algorithm should be used. For real world applications, like logo removal and object removal, image inpainting can also be utilised as a tool. Patch based techniques [2],[3] and Diffusion based methods [1] are two older workshop orders that can be distinguished. While the final method uses appearance copying and pasting to spread the contents of the image from known regions to unknown corridors, the earlier method uses partial discrimination equations and variational styles to spread information from the girding regions to the interiors of the incomplete regions. When dealing with little missing portions, these ways have produced excellent visual results. Still, they're unfit to make substantial structures and objects that are absent from other corridor of the image for huge missing portions. Recent research uses deep learning to resolve this problem, including convolutional neural networks (CNNs) [4] and generative adversarial networks (GANs) [5]. These inpainting techniques can be loosely divided into three groups from the aspect of network design: one-stage networks (i.e., one generator) [6]-[10], progressive networks (one or more than one generators applied repeatedly) [12] and two-stage networks (i.e., two generators) [11] are more examples. In this research, we develop an image reconstruction with inpainting network using this method using the receptive field, a collection of input pixels that are directly coupled to a neuron. [15]. Three things serve as inspiration for our work: (1) The digital image picture inpainting issue is related to the receptive field, and the size of the surrounding areas needed to fill in different types of missing patches varies. (2) Large receptive fields, which the majority of earlier approaches aimed for, would not be the best option, particularly for mending the regional structures and specifics. A broad receptive field, on the other hand, is more likely to contain undesirable completion outcomes, which could be destructive to the inpainting process. (3) Deep neural networks, which have attracted increased interest in image classification and semantic segmentation [16] have an important component known as the receptive field.

The following contributions are made by our work:

- To enhance the inpainted image, our small refining network has a constrained receptive field. This shallow deep network can recover some detail following the partial inpainting stage. missing regions, such as small structures and texture details, in accordance with the nearby small regions and prevent long-distance failed completion results from having an impact.

- We suggest an attention based large refining network with a large receptive field to further enhance the completion outcome. Using large data, this network can significantly enhance visual quality, particularly for massive objects and distant texture patterns. Due to the relatively high quality of the small refinement network's output, the calculation of attention is also more reliable and steady.

- On three well-known public inpainting datasets, CelebA-HQ [17] (Fig.1), and Paris StreetView [19] (Fig.2), our suggested inpainting system performs at the cutting edge.

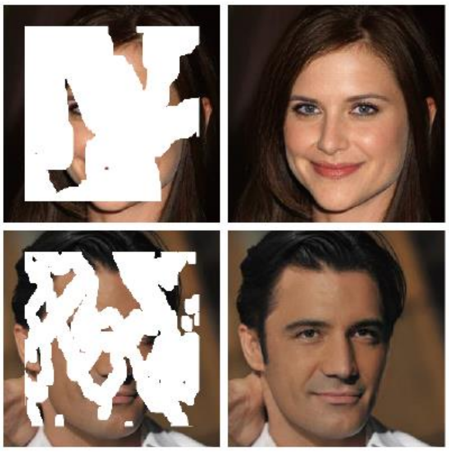

Fig.1 CelebA-HQ sample dataset

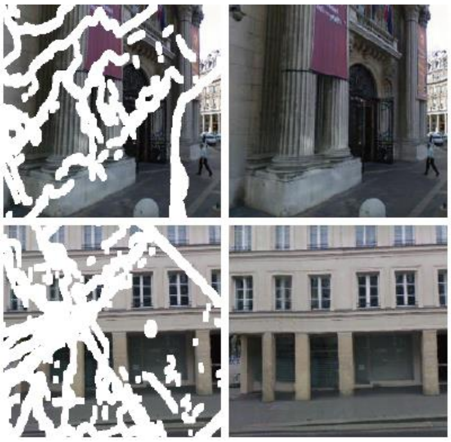

Fig. 2 Paris Street View sample dataset

2 Related Works

There are two types of inpainting methods used up till now are known as classical inpainting methods and deep learning based inpainting methods.

2.1. Classical Inpainting Methods

The traditional methods include two types they are Diffusion based techniques and Patch based techniques.

2.1.1 Diffusion based techniques.

Diffusion based method only focus on the small region in the inpainting process. Its quality is highly influenced by the known data that are selected. Bertalmio et al. were the ones who first used diffusion for image reconstruction. [1] Once the location to be painted has been chosen, the method they provide doesn't need any human input. Without the user selects "what to place where," the algorithm may concurrently fill areas surrounded by vaus backdrops. The technique utilises straight lines to connect pixels with the same grey, therefore the angle at which the level lines arrive at the boundary of the inpainted zone and are not (fully) retained. Their technique resolves these issues, which will be shown later in this study. The limitation of this work is it does not efficiently reproduce the texture.

2.1.2 Patch based techniques.

Patch level similarity is a key component of patch-based approaches, which are used to transfer appearance information from the background areas to the missing components. Barnes and others, [2] introduced a randomised nearest-neighbour patch technique, called Patch Match, which is commonly used in editing tools, to decrease the cost and time for patch matching. According to this algorithm, certain good patches are discovered by random sampling and the imagery's inherent coherence enables us to quickly spread these matches to adjacent regions. The limitation of this process is only focusing of the nearest neighbour fields and it's failed to works with far away fields. Planar perspective and translational regularity are two examples of mid-level restrictions used by Jia-Bin Huang et al. [3] to direct the low level completion process. They begin by identifying the various planes and figuring out where in the scene they are supported spatially. They use SIFT feature matching to ascertain the translational regularity within each plane.

2.2 Image Inpainting Based on Deep Learning methods.

There are three categories of deep learning-based image inpainting techniques exist: one-stage, two-stage, and progressive techniques.

2.2.1 One Stage Methods

Goodfellow et al. [4] created a system for assessing generative models using an adversarial process in early approaches. This network comprises of a generative model for data distribution and a discriminative model that estimates the likelihood that a sample was drawn from training data as opposed to generative model output. Maximizing the likelihood that the discriminative model will be incorrect is the primary objective of the generative model. Pathak et al.'s[5] training used context encoders, and they experimented with adversarial loss and pixel-wise reconstruction loss. In this paper, they suggested a pixel prediction-based context-based unsupervised visual feature learning system. They worked on three distinct sections of the mask (missing regions). i) Central region This is effective for painting tasks because the network learns low-level characteristics that grab onto the edge of the core mask, but the features this area learnt weren't all-encompassing. ii)Random block-In this type instead of giving large single area as a mask region they removed a number of smaller possibly overlapping masks that covers up to one fourth of an image. iii)Random Region- In this type they remove the shapes from ground truth image. Convolutional filter responses are used in deep learning-based image reconstruction techniques in order to apply CN over damaged images that contain both valid and invalid pixels (masked regions). This causes a blur effect and colour disparity. Iizuka et al.[6] proposed the technique, The Context Encoder (CE) methodology, which makes use of a Convolutional ACM Transactions on Graphics. The CE approach, which was driven by feature learning. This enhances the outcomes of the visual quality. Models can be trained to produce realistic image completeness. The key benefit of this method above conventional ones like Patch Match is that it can create unique objects that don't already exist in the image. As a result of this method's explicit training to be both small and largely consistent, images are completed in a far more natural way. Large holes cannot be filled in with this model's spatial support, which is a drawback. Liu et al.[7] proposed partial convolution. It performs the normalization of the output to adjust for the fraction of missing data. Here the convolution is masked and renormalized to work only on valid pixels. They also performed a function that automatically generates the updated mask for the next layer. Their model performs well for irregular mask. Zeng et al.[8] propose the Pyramid-context ENcoder Network (PEN-Net) for deep generative models for image inpainting. They suggest a pyramid-context encoder that gradually transfers the learnt attention to the preceding low-level feature map while learning area affinity via attention from a high-level semantic feature map. In early works the structures and textures were recovered in a step by step manner they are not considered them as a whole. This leads to a insufficient usage of encoder. So Hongyu Liu et al.[9] proposed a mutual encoder-decoder for the recovery of both structures and textures. They use CNN features from the deep and shallow layers. The deep layers used for structure recovery and shallow layers used for texture recovery. Hui et al.[21] proposed this model to overcome the problem of unreasonable structure or blurriness. In this paper one-stage model that uses dense combinations of dilated convolutions were implemented. For the training of the generator, they created self-guided regression loss for concentrating on certain areas. To maintain uniformity between small and large contents, they additionally use a discriminator with small and large branches.

2.2.2 Two Stage Methods

Yu et al.'s[10] proposed a model that primarily uses both conventional and deep learning-based inpainting techniques, which, in addition to directly using the surrounding image attributes as references at the time of the network training to produce better predictions, may synthesise original image structures. Nazeri et al.[11] proposed a structure guided image inpainting network with edge prediction. This model was comprised of two stages of work that is edge generator first their model forecasts the structure of the image's missing region later this structure will be generated into second stage for image inpainting process. In this study, Wu et al. [23] combined an inpainting network with a small binary pattern (LBP) network to create a novel end-to-end two stage generative model. U-Net structure, which forecasts the structural information of missing areas, was utilised by the LBP network. The inpainting network uses this as guidance to fill in the missing pixels more effectively.

2.2.3 Progressive methods

Guo et al. [13] suggests the full-resolution residual network (FRRN), which has been demonstrated to be successful for progressive picture inpainting, to fill irregular holes. Due to the fact that each residual block only aims to reconstruct a specific region and since stacking residual blocks is ideal for this progressive inpainting method, a residual architecture for progressive picture inpainting may also be advantageous. Li et al. [14] build a Knowledge Consistent Recurrent Feature Reasoning (RFR) network that essentially consists of Attention (KCA) and plug and play modules for recurrent feature reasoning. The internal content constraints are gradually tightened, and the model can generate semantically precise outputs.

3 Methodology

3.1 Partial Inpainting Network

A Partial Inpainting Network's objective is to take an input image that is damaged or incomplete and turn it into an output image that is complete and aesthetically realistic. In order to determine what the missing portions of the image should look like, the network analyses the parts of the image that are viewable. In order for a Partial Inpainting Network to learn how to fill in missing areas in a variety of cases, it is typically trained on a huge dataset of images that have been intentionally damaged or made incomplete. The network is often made up of several layers of neurons, each of which processes the input image and refines the output to get a more precise and detailed result. Our Partial Inpainting Network (NetP) is built using an encoder-decoder structure(UNet) as its foundation. Eight processes make up the network, and they cycle between down-sampling and up-sampling the image. With successive processing of the input image, the network is able to gradually improve its output. The network can keep fine-grained features and avoid information loss during the down-sampling phase with the help of skip connection. The information is sent from the encoder to the decoder via long skip links in order to retrieve the data that was lost during down sampling. A large receptive field helps the entire structure come together. The network will receive an input picture I_{in} and a binary mask M specifying the incomplete areas, where 0 indicates a valid pixel and 1 indicates a missing pixel. Our Partial Inpainting Network NetP produces an inpainted image called I_{out}^{P}. The training goal of NetP consists of an adversarial loss and a pixel-wise reconstruction loss. In our approach, we rebuild pixels using the weighted L1 loss,

$$L_{valid}^{P}=\frac{1}{sum(1-M)} \parallel (I_{out}^{P}-I_{gt})\odot(1-M) \parallel 1 \quad (1)$$

$$L^P_{hole}=\frac{1}{sum(M)} \parallel (I^P_{out} - I_{gt})\odot(M) \parallel 1 \qquad (2)$$

where I_{gt} stands for the ground-truth picture, sum(M) and the number of non-zero components are represented as sum and stands for the element-wise product operation (M). the following is then used to express the pixel-wise reconstruction loss:

$$L^P_r= L^P_{valid} +\lambda_h \cdot \mathrm{L}^{\mathrm{P}}_{\mathrm{hole}} \qquad (3)$$

Balancing factor is represented as λ_h.

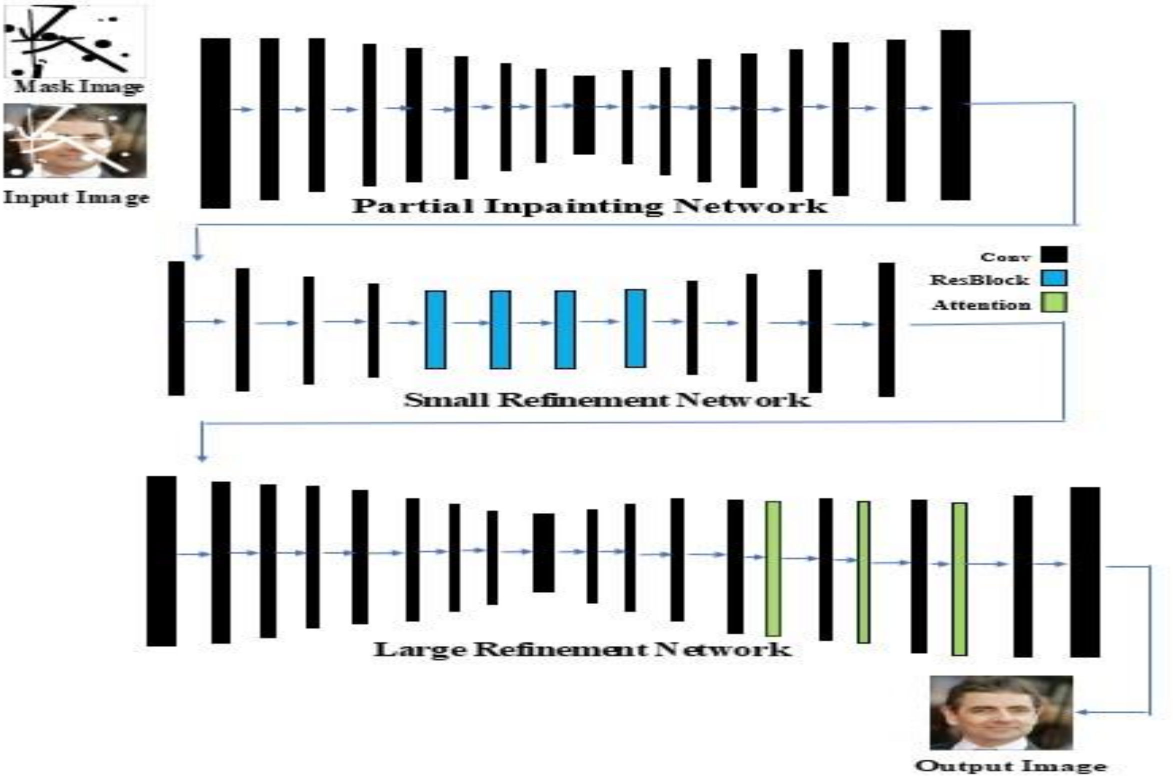

Fig 3 Proposed Network Architecture

Figure.3.The network design of our suggested technique. Four blue bricks are used to depict the leftover block. The Attention modules are represented by 3 green blocks in a large refining network.

3.2 Small Refinement Network

A particular kind of neural network called a small inpainting network is created specifically for the task to fill the missing or damaged areas of a picture. The small inpainting network was created with the particular purpose of emphasizing on nearby areas of an image and producing high-quality results. The small inpainting network concentrates on small portions of the image at a time, as compared to a Partial Inpainting Network, which works on the entire image at once. Convolutional neural networks (CNNs), which are applied to extract and process image data at various resolutions, are frequently used in the small inpainting network. The network makes advantage of these properties to generate new pixels and complete any gaps in the image. The ability of small inpainting networks to catch fine-grained information and provide results that are visually consistent with the surrounding regions allows them to produce extremely high-quality results for small regions of a image.

A shallow deep network is created for the small refining. Four residual blocks, two up-sampling operations, and two down-sampling operations forms the small refinement network (NetS). The shallow nature of the network causes it to have a narrow receptive field , which it uses to process the small region of the partial inpainted network result using the a sliding window design technique. This design allows for some empty missing areas, such as minor buildings and textures, to be filled appropriately using adjacent small data, independent of more distant and failed filling contents. And Then, we have used extra residual blocks to see whether they could progressively increase the receptive field, albeit the inpainted findings only indicate marginal gains.
The weighted reconstruction loss $\cdot L_r^S$ is the first objective for NetS's training. As a smoothing penalty, the total variation (TV) loss is performed.

In order to reduce the discrepancy between the characteristics extracted from the target picture image and those from the output image, a form of loss function known as perceptual loss [24] is utilised in deep learning models. Perceptual loss is frequently employed in tasks like picture style can transfer the image super resolution, where the output image must maintain the style or level of detail of the input image in addition to being visually comparable to it. In comparison to employing a more straightforward loss function, such as mean squared error, the output image produced by applying a perceptual loss function can be more aesthetically pleasing and closer to the original image. In image processing and computer vision, the difference in style between two pictures is referred to as "style loss." It is used as part of the neural style transfer method, which uses a neural network to transfer the style of one picture to another. The style loss is often evaluated as the mean squared error between the feature maps of the two pictures after each image has been processed using a pre-trained convolutional neural network (CNN) and its feature maps have been extracted at various levels. The neural network is used by minimising the style loss between the input picture and the style image, the network may create a new image that combines the content of the input image with the style of the style image. Similar to many prior efforts [8][10] and [14] style loss [25] defined on the VGG-16 [26] (pre-trained on ImageNet [27]) are also employed to improve the recovery of the structural and textual information. In contrast to the pixel-wise reconstruction loss and TV loss outlined above, which are calculated in the pixel space, these two losses are computed in the feature space. Total Variation (TV) loss, a type of regularisation approach used in image processing, reduces noise while keeping the picture's edges and features.TV loss treats the image as a 2D grid of pixels, and it calculates the overall variation of the image as the sum of the absolute differences between adjacent pixels. As a result, a measurement of the image's total variation is produced. This measurement can be used to regularise the image and lower its noise level. In conclusion, the small refinement network's objective is:

$$L_S = L_{valid}^S + \lambda_h \cdot L_{hole}^S + \lambda_{tv} \cdot L_{tv}^S + \lambda_{per} \cdot L_{per}^S + \lambda_{sty} \cdot L_{sty}^S$$

3.3 Large Refinement Network

Once the missing regions have been roughly filled in, large information is used to refine the estimation and generate a more plausible completion. Several methods, including deep learning-based methods, partial differential equations for picture inpainting, and texture creation, can be used to accomplish this. We offer an attention-based large refinement network(NetL) that uses an attention method and a wide receptive field to maximise the amount of data that a neuron can take in. Since our Partial Inpainting Network already has enough receptive field to cover the entire image, we merely use the NetP network architecture. Small refinement networks can deliver completion results that are largely accurate. As a result, the large refinement network's attention computation tends to be more reliable and stable. Existing publications [11] and [14] make substantial use of the attention scheme to explain connections between contextual information and the missing areas like symmetry and repeating designs, for example. In our research, the simple self-attention approach is used. [11], [14], while a more sophisticated attention technique may potentially be applied in our framework. UNet and UNet++ are both popular neural network architectures used for semantic segmentation tasks, but UNet++ is an extension of the original UNet architecture with additional skip connections. UNet++ replaces the single skip connection between each encoder and decoder block in UNet with a series of nested skip connections that connect all the previous encoder and decoder blocks. This hierarchical architecture allows UNet++ to capture more contextual information and improve the accuracy of semantic segmentation.so we also tried UNet++ architecture in the place of UNet. It increases the accuracy a bit more.

4 Experiments

In this part, we first go through the experimental settings, datasets, implementation specifics, and Metrics calculation.

4.1 Experimental Settings

4.1.1 Datasets

Two open datasets that are frequently used to gauge the effectiveness of picture inpainting activities serve as the basis for our investigations.

• CelebA-HQ dataset [17] :The CelebA [17] high-quality version contains 30,000 facial photos. The test split that we provide is 0.3, which is 30%. Hence, 21,000 photos are utilised for training, whereas 9,000 images are used for testing purpose.

•Paris Street View dataset [19]: Paris Street View dataset Imagery taken at the street level makes up this dataset. The initial setup is kept, with 13000 photos used as the training set and 1000 photos used as the test set.

We build the irregular masks using a few fundamental procedures on the QD-IMD (quick draw irregular mask dataset) to train the networks. We employ Liu et al[8] .'s irregular mask data as the testing masks to assess the trained models in accordance with the accepted methodologies. These masks are produced in [8] using random dilation, rotation, and cropping.

4.1.2 Implementation Specifics

Our SLRNet is implemented in tensorflow. With the aid of TensorFlow, users may create deep neural networks for use in processes like image identification, categorization, and restoration as well as natural language processing. All of the photos and masks in our project are 256X256 pixels in size.

4.2 Metrics

4.2.1 Quantitative comparisons

There are number of standard measures for the image inpainting job as the assessment criteria, including: The structural similarity index (SSIM), Frechet inception distance (FID), learning perceptual image patch similarity (LPIPS), peak signal-to-noise ratio (PSNR), and L1 error are some examples of these metrics. The first three metrics are based on low-level pixel values, and the latter two metrics are related to high-level visual perception. The performance of our suggested method is the finest of all inpainting techniques.

We calculated the first three metrics for our project. L1 error, also known as the Mean absolute error (MAE) is a statistic that assesses how accurately a model predicts the future. It is determined by taking the absolute difference between the actual values and the projected values, average these absolute differences over all samples in the dataset, and then calculating the difference in percentage terms. The PSNR (Peak Signal-to-Noise Ratio)gauges how well an image or video signal has been rebuilt. Quantifying the difference between an original image and one that has been compressed or warped is a typical practise in image processing.

The SSIM index compares two images based on how similar they appear to the human visual system in terms of luminance, contrast, and structure. By dividing each image into smaller image blocks and comparing the similarities between matching blocks in each image, the SSIM index compares two photos.Luminance, contrast, and structure are the three factors used to generate the SSIM index.A value between -1 and 1 is commonly used to represent the SSIM index, with values closer to 1 denoting greater similarity between the two images. Perfect similarity is represented by a value of 1, whereas perfect dissimilarity is represented by a value of -1

Fig. 4 Quantitative comparisons on Celeb-A dataset

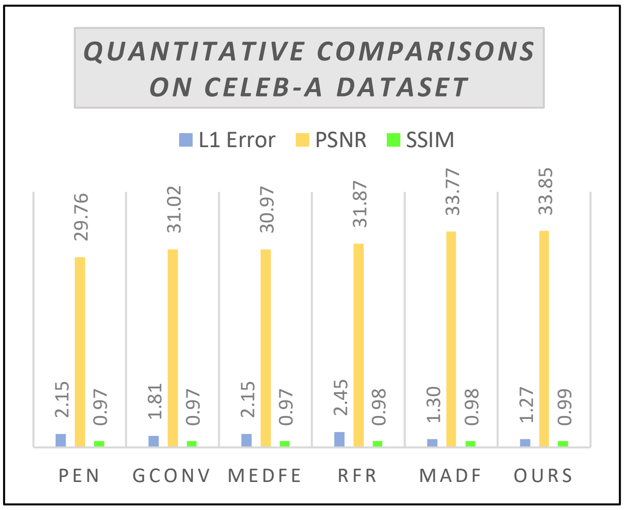

Fig 4 Illustrates the Quantitative comparisons on Celeb-A dataset for the metrices L1 Error, PSNR, SSIM of mask(10-20%)

Fig. 5 Quantitative comparisons on Paris street view dataset

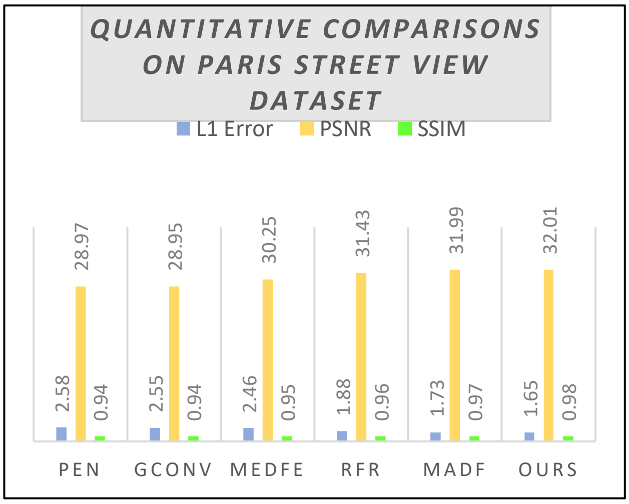

Fig 5 Illustrates the Quantitative comparisons on Paris street view dataset for the metrices L1 Error, PSNR, SSIM of mask(10-20%)

Table 1

Quantitative comparisons on Celeb-A dataset. ǂ - Lower is better. Ͳ – Higher is better.

	Masks	1-10%	10-20%	20-30%	30-40%	40-50%	50-60%
L1(%) ǂ	**PEN**	**0.80**	**2.15**	**3.88**	**5.83**	**8.02**	**11.77**
	GConv	**0.65**	**1.81**	**3.41**	**5.33**	**7.53**	**12.05**
	MEDFE	**1.02**	**2.15**	**3.68**	**5.51**	**7.65**	**11.67**
	RFR	**1.59**	**2.47**	**3.58**	**4.90**	**6.44**	**9.47**
	MADF	**0.47**	**1.30**	**2.40**	**3.72**	**5.26**	**8.43**
	Ours	**0.45**	**1.27**	**2.38**	**3.72**	**5.25**	**8.40**
PSNR Ͳ	**PEN**	**35.34**	**29.76**	**26.79**	**24.70**	**23.06**	**20.85**
	GConv	**37.14**	**31.02**	**27.57**	**25.03**	**23.10**	**20.22**
	MEDFE	**36.13**	**30.97**	**27.75**	**25.35**	**23.47**	**20.85**
	RFR	**36.39**	**31.87**	**29.07**	**26.87**	**25.09**	**22.51**
	MADF	**39.68**	**33.77**	**30.42**	**27.95**	**25.99**	**23.07**
	Ours	**40.01**	**33.85**	**30.43**	**27.97**	**26.01**	**23.11**
SSIM Ͳ	**PEN**	**0.988**	**0.965**	**0.933**	**0.894**	**0.849**	**0.764**
	GConv	**0.991**	**0.971**	**0.941**	**0.902**	**0.856**	**0.750**
	MEDFE	**0.990**	**0.971**	**0.943**	**0.908**	**0.865**	**0.775**
	RFR	**0.991**	**0976**	**0.957**	**0.932**	**0.902**	**0.834**
	MADF	**0.994**	**0.983**	**0.967**	**0.945**	**0.917**	**0.848**
	Ours	**0.995**	**0.990**	**0.967**	**0.945**	**0.917**	**0.849**

Table 2

Quantitative comparisons on Paris Street View dataset. ǂ - Lower is better. Ͳ – Higher is better.

	Masks	1-10%	10-20%	20-30%	30-40%	40-50%	50-60%
L1(%) ǂ	**PEN**	**0.97**	**2.58**	**4.65**	**6.84**	**9.35**	**13.00**
	GConv	**0.92**	**2.55**	**4.67**	**6.99**	**9.58**	**14.19**
	MEDFE	**1.15**	**2.46**	**4.24**	**6.25**	**8.63**	**12.73**
	RFR	**0.71**	**1.88**	**3.38**	**5.04**	**6.95**	**10.28**
	MADF	**0.64**	**1.73**	**3.19**	**4.86**	**6.79**	**10.33**
	Ours	**0.58**	**1.65**	**3.06**	**4.78**	**6.65**	**9.99**
PSNR Ͳ	**PEN**	**34.25**	**28.97**	**26.03**	**24.12**	**22.56**	**20.72**
	GConv	**34.72**	**28.95**	**25.73**	**23.62**	**21.95**	**19.59**
	MEDFE	**35.12**	**30.25**	**27.03**	**24.91**	**23.12**	**20.76**
	RFR	**36.81**	**31.43**	**28.39**	**26.30**	**24.60**	**22.27**
	MADF	**37.64**	**31.99**	**28.71**	**26.44**	**24.65**	**22.14**
	Ours	**37.65**	**32.01**	**29.30**	**27.01**	**25.01**	**22.17**
SSIM Ͳ	**PEN**	**0.979**	**0.939**	**0.884**	**0.821**	**0.745**	**0.624**
	GConv	**0.980**	**0.940**	**0.885**	**0.825**	**0.757**	**0.629**
	MEDFE	**0.984**	**0.954**	**0.909**	**0.854**	**0.787**	**0.660**
	RFR	**0.987**	**0.962**	**0.928**	**0.886**	**0.836**	**0.733**
	MADF	**0.989**	**0.966**	**0.933**	**0.892**	**0.841**	**0.732**
	Ours	**0.990**	**0.981**	**0.938**	**0.897**	**0.845**	**0.734**

4.2.2 Qualitative comparisons

Two clusters, each with three or four rows, stand for CelebA-HQ and Paris, respectively. Our observations of PEN and G-Conv's quality are in line with the quantitative results. Our method is more successful in repairing facial features such the lips, nose, and eyes when comparing the results of face completeness. As it is based on the relatively strong completion result of the small refinement network, our attention computation in the large refinement network is more dependable and stable than G-Conv and PEN.In addition, our approach can better restore the structures and details when compared to current inpainting techniques. The small refinement network is used to complete the local structures and textures since it has residual block to increase the receptive field size. The next step is to improve the remaining missing parts of a picture and increase its quality using a vast refinement network using an attention scheme mechanism.

5 Results

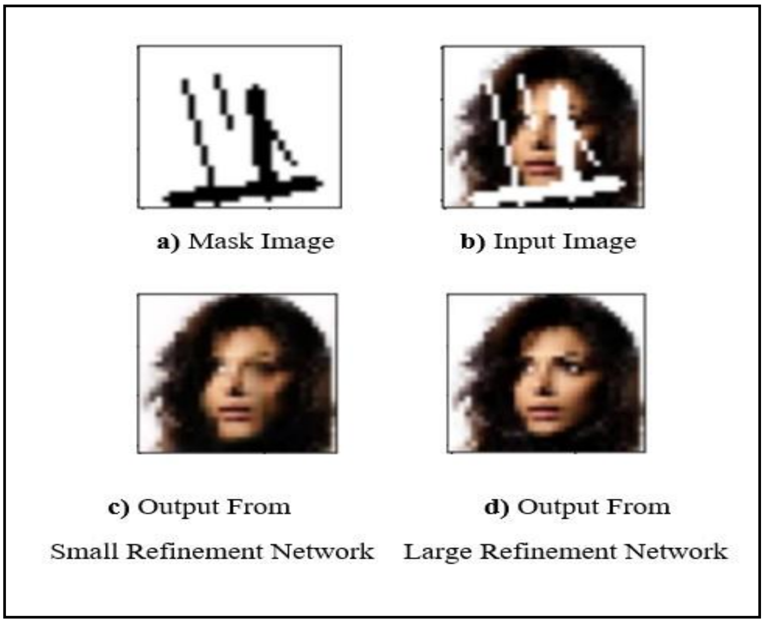

Fig. 6 Quantitative comparison on Celeb-A dataset

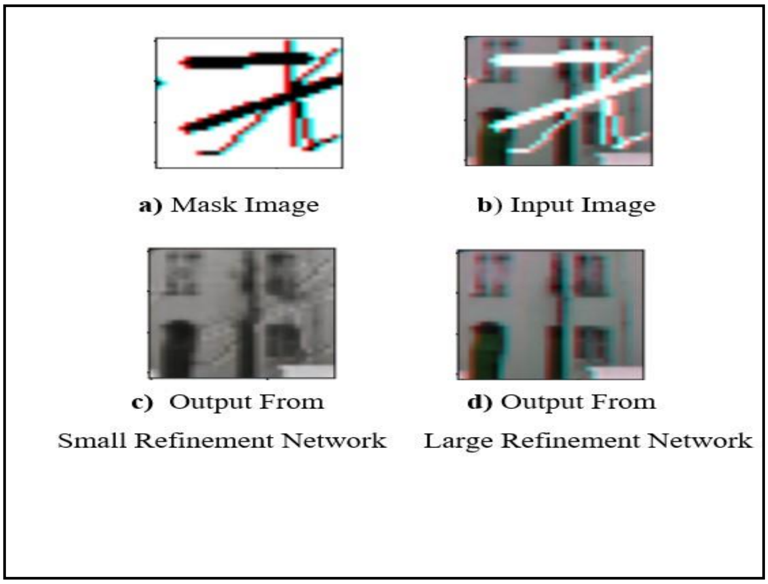

Fig. 7 Quantitative comparison on Paris Street view dataset

6 Conclusion and Future work

For image inpainting, we presented a three-stage network that takes nearby pixels into account. A Partial Inpainting Network with a large receptive field is used to complete the whole structure as well as part of the texture details (set of input pixels).

A small refining network with a residual block and a limited receptive field serves the twin purposes of removing visual impacts that are strongly related to the small region and protecting against the detrimental effects of distant and failed filling contents. It is recommended that an attention-based large refinement network with a large receptive pitch be used to combine the large information with the more accurate attention computation to further improve the visual quality of the inpainted results.

Our project consists of three networks among those we use small and large receptive field based inpainting network in a separate model.it leads to more storage of data so inorder to resolve it use residual block(used in Small Refinement Network) and attention scheme (used in Large Refinement Network) in a single network called small and Large inpainting network.

Data availability We have used two publicly available datasets [17][19].

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,”in Proc.ACM SIGGRAPH, 2000, pp. 417–424.

[2] C. Barnes, E. Shecht man, A. Finkelstein, and D. Goldman, “PatchMatch:A randomized correspondence algorithm for structural image editing,” ACM Trans. Graph., vol. 28, no. 3, p. 24, 2009.

[3] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, “Image completion using planar structure guidance,” ACM Trans. Graph., vol. 33, no. 4, pp. 1–10,Jul. 2014.

[4] O. Ronneberger, P. Fischer, and T.Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. Cham,Switzerland:Springer, 2015, pp. 234–241.

[5] I. Goodfellow et al., “Generative adversarial nets,” in Proc. Adv. Neural Inform.Process. Syst., 2014, pp. 2672–2680.

[6] D. Pathak, P. Krahenbuhl, J. Donahue,T. Darrell, and A.A. Efros “Contextencoders: Feature learning by inpainting,” in Proc. IEEE Conf.Comput. Vis. Pattern Recognit.(CVPR), Jun. 2016, pp. 2536–2544.

[7] S. Iizuka, E. Simo-Serra, and H.Ishikawa, “Largely and smallly consistent image completion,” ACM Trans. Graph., vol. 36, no. 4, pp. 1–14, 2017.

[8] 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 Proc.Eur. Conf. Comput. Vis., Sep. 2018, pp. 85–100.

[9] Y. Zeng, J. Fu, H. Chao, and B.Guo, “Learning pyramid-context encoder network for high-quality image inpainting,” in Proc. IEEE/CVF Conf.Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 1486–1494.

[10] H. Liu, B. Jiang, Y. Song, W.Huang, and C. Yang, “Rethinking image inpainting via a mutual encoder-decoder with feature equalizations,” in Proc. Eur. Conf. Comput. Vis., 2020,pp. 725–741

[11] J. Yu, Z. Lin, J. Yang, X. Shen, X.Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” in Proc. IEEE/CVF Conf.Comput. Vis. Pattern Recognition.,Jun. 2018, pp. 5505–5514.

[12] H. Zhang, Z. Hu, C. Luo, W. Zuo,and M. Wang, “Semantic image inpainting with

progressive generative networks,” in Proc. 26th ACM Int. Conf. Multimedia, Oct. 2018, pp.1939–1947.

[13] Z. Guo, Z. Chen, T. Yu, J. Chen,and S. Liu, “Progressive image inpainting with full-resolution residual network,” in Proc. 27th ACM Int. Conf. Multimedia, Oct. 2019, pp. 2496–2504.

[14] J. Li, N. Wang, L. Zhang, B. Du, and D. Tao, “Recurrent feature reasoning for image inpainting,” in Proc. IEEE/CVF Conf. Comput. Vis.Pattern Recognit. (CVPR), Jun.2020, pp. 7757–7765.

[15] J. Long, N. Zhang, and T. Darrell, “Do convnets learn correspondence?” in Proc. Adv. Neural Inform. Process. Syst.,2014, pp. 1601–1609.

[16] W. Luo, Y. Li, R. Urtasun, and R.Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” in Proc. Adv.Neural Inform. Process. Syst., 2016, pp. 4905–4913.

[17] T. Karras, T. Aila, S. Laine, and J.Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation,” in Proc. Int. Conf. Learn. Represent., 2018, pp. 1-26.

[18] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places:A 10 million image database for scene recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 6, pp. 1452–1464, Jun. 2018.

[19] C. Doersch, S. Singh, A. Gupta, J.Sivic, and A. A. Efros, “What makes Paris look like Paris?” ACM Trans. Graph., vol. 31, no. 4,pp. 101:1–101:9, 2012

[20] O. Elharrouss, N. Almaadeed, S. Al Maadeed, and Y. Akbari,“Image inpainting
A review,” *Neural Process. Lett.,* vol. 51, no. 2,
pp. 2007–2028, 2019.

[21] Z. Hui, J. Li, X. Wang, and X. Gao, “Image fine-grained inpainting,” 2020, *arXiv:2002.02609.*

[22] M. Zhu *et al.*, “Image inpainting by end-to-end cascaded refinement with mask awareness,” *IEEE Trans. Image Process.*, vol. 30, pp. 4855–4866, 2021.

[23] H. Wu, J. Zhou, and Y. Li, “Deep generative model for image inpainting with small binary pattern learning and spatial attention,” 2020, *arXiv:2009.01031.*

[24] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Proc. Eur. Conf. Comput.Vis.*, Oct. 2016, pp. 694–711.

[25] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in *Proc. IEEE Conf. Comput.Vis.Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[26] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet:A large-scale hierarchical image database,” in *Proc. IEEE*