Research Article

Application of Image Inpainting in Drug and Alcohol Addiction Research Using Repeated Convolution Method

Jini P* and **Rajkumar KK**

Department of Science and Information Technology, Kannur University, India

*Address Correspondence to Jini P, E-mail: jininandanam@yahoo.com

Abstract

Image inpainting is a promising but challenging approach that fills in huge free-form empty areas in images. Most of the recent papers concentrate on splitting masked image into 2 matrices of valid and invalid elements which makes the system more complex. This paper proposes a novel algorithm named ReConv which uses a repeated standard convolution operation which treats valid and invalid elements of an image in the same manner. The outcomes of our suggested method, ReConv, shows that, in comparison to earlier approaches, our system produces outputs that are more adaptable with good quality for real world applications. In the context of drug and alcohol addiction treatment and research, this technology offers several unique and emerging applications like Therapeutic Visual Stimuli Modification. Inpainting techniques can fill in missing data in addiction-related images, such as damaged MRI scans or incomplete survey responses, enhancing the predictive capacity of machine learning models used in addiction research. An extensive comparison study on 2 types of datasets validates our method. The effectiveness of the suggested strategy was evaluated using different measures such as PSNR, SSIM and FID. The results show that our recommended approach excels in performance compared to the existing modern methods.

Keywords: Image inpainting; Image restoration; Auto encoder; Repeated convolution

Introduction

Image inpainting is a highly utilized subject matter in the past few years [1]. Image inpainting process contributes a fantastic appearance to an image in this manner that an average person is not able to recognize that the image has undergone certain alterations [1]. It can restore missed portion of an image by using accessible data of exactly from that image itself. Image inpainting is commonly employed in image processing applications like elimination of superimposed text like dates, captions, as well as stamps in different images. The purpose of inpainting an image is to fill in vacant spaces in images with logical information and in addition, it manages further computer vision tasks like object detection, object identification, and semantic segmentation [2]. Despite the many benefits of this approach, retrieving suitable, rich, and distinctive textures for broad unstructured missing portions in high-quality photographs presents major challenges for image inpainting. Although in traditional methods there are numerous substitute techniques for image restoration, none of them use any effective digital image inpainting methods. First image inpainting technique was emerged in 2000 known as diffusion based technique or Partial Differential Equation based method (PDE) [1]. It also known as pixel oriented method. This method uses mathematical approaches such as Partial Differential Equations (PDEs) in order to spread the available data into the missing region. PDE works by diffusing image pixels from the target region's edge to its interior by propagating information from the border. Contextual information is moved from borders into holes in the direction of the isophotes by this method [1,2].

Even though, this method is very effective in filling small regions, it tends to introduce some blurry effect while filling large textured regions. This occurs because PDE solutions frequently require the boundary and initial conditions to be provided [2]. The primary drawbacks of this model are that it performs inadequate when reconstructing large textured areas because the propagation process introduces blurring artefacts and there is insufficient explicit management of the edge pixels. The answers might not be evident since it might be challenging to get exact boundary conditions in real-world situations. Additionally, solving PDE equations requires additional computing time [3].

Another well-known image inpainting technique derived after PDE is Exemplar based method or Patch-based

method. In Patch-based approach, patches are filled in appropriate places by giving priority to each patch. Exemplar Based approach fails to reconstruct structured regions and is primarily focused on filling huge textured regions, while being faster and more efficient than PDE. The patch's size has an effect on how inpainting turns out. There isn't yet a simple method for figuring out the patch's size automatically [4]. The best-fitting patches are continuously found using patch match algorithm to close the gaps. Although this approach typically yields smooth results, its ability to understand visual semantics is limited by the given image information. This algorithm concentrates on rectangular-shaped holes, which are frequently considered to be in the middle of image and the search is incredibly poor and prone to inaccurate results [5]. This is a main drawback of Exemplar based method and eventually decrease the usefulness of these models in the application [6]. A new inpainting algorithm called Fast Marching Method [7] the basis for obtaining FMM is the propagation of an image smoothness estimator down the image gradient. It is almost similar to PDE method [1]. By using a weighted average over an established region around the pixel, this approach estimates the smoothness of the image. They transmit image information using Fast Marching Method (FMM) outlined in, treating the damaged regions as level sets [7]. Anyway, due to the lack of specific methods for inpainting edge regions, these rapid strategies are not appropriate for filling large hole regions [8].

Another popularly used method for image inpainting is using image interpolation in image inpainting it ignores the size of the damaged area [9]. Since the techniques mentioned above concentrate on the damaged region's size, interpolation technique treats small and large region in a similar manner. This method uses information from the surrounding pixels to complete damaged areas of an image. Some popular image interpolation methods are nearest neighbour interpolation, Bilinear interpolation, Bicubic interpolation and Kriging interpolation [10]. All these techniques have its own corresponding equations and the computational complexity of these equations makes it really challenging in image inpainting applications [11].

However, these conventional image inpainting techniques frequently fall short when the missing portion is significant or complicated since they are unable to extract deeper elements from the original image, such as higher level semantic, texture, and other features [12]. Researchers can utilize deep learning models to solve computer vision issues. We introduce a novel auto encoder-based method in this paper to deal with the previously noted problems with high-resolution image restoration.

Previously published papers consider valid and invalid elements of the masked image as differently [13,14]. This frequently results in distortions like color disparity and blurriness. In our work we consider both these type of elements in the same manner and this is the highlighted simplicity of our work. For irregular masks, our model beats existing techniques. Targeting for advanced image

inpainting, we suggest for using repeated convolution method for fine-tuning of the result. To support future efforts at developing and testing inpainting models, we propose to create a significant dataset of irregular masks that will be released to the public. We validate our methodology through qualitative and quantitative comparisons using standard measurements available in the literature [15]. We perform in-depth analyses on 2 type of dataset named CIFAR-10 and CelebA datasets [16].

By harnessing image inpainting, professionals can create powerful, personalized tools for therapy, improve the accuracy of medical research, and raise public awareness, offering fresh possibilities in the fight against drug and alcohol addiction. Machine learning models for diagnostic purposes rely on high-quality medical images. Inpainting generates synthetic yet realistic data by completing incomplete scans, thus augmenting datasets used for AI model training. Inpainting can help reconstruct and predict how certain tissues or organs might change over time, which is especially useful in longitudinal studies for diseases like Alzheimer's, cancer, or multiple sclerosis.

Image inpainting in drug and alcohol addiction treatment and research

Image inpainting, a technique used in artificial intelligence and computer vision, refers to filling in or reconstructing missing, corrupted, or intentionally removed parts of an image. In the context of drug and alcohol addiction treatment and research, this technology offers several unique and emerging applications.

Therapeutic visual stimuli modification

Trigger desensitization: Individuals recovering from addiction often experience cravings triggered by visual cues (e.g., alcohol bottles or drug paraphernalia). Inpainting can be used to modify photographs to remove or replace these cues, helping patients practice desensitization to real-world stimuli in a controlled environment.

Before-and-after simulations: Inpainting can generate visual transformations showing improvements after lifestyle changes, such as depicting the reversal of physical damage caused by substance abuse. This can motivate individuals during recovery.

Neurofeedback and Virtual Reality (VR) applications

VR-based therapy: In virtual reality environments, inpainting can dynamically alter visual scenes, removing elements that might trigger addiction responses or creating new positive scenarios for behavioral conditioning.

Customized therapy sessions: Neurofeedback devices that detect brain responses to visual cues can use inpainting to modify stimuli in real-time, helping therapists monitor and adjust treatment.

Privacy-preserving research

Anonymized image data: Researchers studying addiction behaviors (e.g., facial cues associated with substance abuse) can use inpainting to remove identifying features

from participant photos while retaining other critical data. This preserves privacy without losing the integrity of research datasets.

Public awareness campaigns

Interactive campaigns: Campaigns focused on addiction prevention can use inpainting to create striking visuals. For instance, an image might show the difference between a person's face before and after addiction recovery, with inpainting providing realistic "healing" simulations to foster public empathy and awareness.

Addiction biomarkers identification

Augmenting medical imaging: Inpainting can be applied in medical imaging studies (e.g., brain MRIs) to reconstruct damaged areas or fill in corrupted scans. This is especially useful when studying brain regions affected by substance abuse. Additionally, AI-enhanced image completion helps researchers observe patterns that may correlate with addiction behaviors.

The remaining portions of this manuscript are divided into the following segments. Section II examines appropriate literature, Section III presents the suggested approach, Section IV discuss the outcomes of the experiment and Section V ends with conclusion.

The main aim behind image restoration is to logically complete spaces remained by damaged portions of images [1]. Image inpainting remains a subject of active research because of its tremendous advantages for image editing features like object removal and image restoration which are very useful [2-4]. The 2 categories of current methodologies are methods depending on learning and methods depending on non-learning.

For completing a particular gap in an image, artists in non-learning image inpainting employs pixels from surrounding areas to complete the damaged component of the image. These techniques perform effectively for inpainting backgrounds in photographs; yet, there are some scenarios where they fail, such as when the surrounding areas lack the essential data to fill in the gaps [4].

While learning-based image inpainting methods concentrate on predicting the missing parts of a damaged image, we will consider a convolution neural network architecture to make predictions of damaged image that are both visually pleasing and functionally effective. We discuss these approaches in detail below.

Image inpainting based on non-learning

Many works on conventional techniques based on non-learning were published in earlier years [1-4]. These approaches do not need training time as in the case of traditional approaches. Traditional non-learning alternatives typically involve old computer vision and image processing approaches. Some of the common approaches are PDE, Exemplar based, and interpolation based [1,3,4,7,9]. Initially, a patch needs to be located in an exemplar-based or patch-based method. After determining which patch best

matches the missing region, the missing pixels are finally calculated using the best matching patch [4]. In these techniques missing pixels are identified by employing multiple neighbouring embedded techniques. The ideal patch size, offset, filling order, and matching algorithms for patch-based algorithms have all been well studied. For instance, the exemplar-based texture generation technique uses picture isophotes and confidence values to determine the target inpainting region's filling order [4,5].

This method uses copies of neighboring pixel patches to fill up the blank spaces [4-6]. Fast Marching Method, a pixel-based technique used by Telea, represented as the pixel data in adjacent unknown areas along with the brightness of an image [7]. It's an easy and quick method of repairing small, homogeneous areas. But this also fails in filling large non homogeneous regions. Ghayoumi et al. (2014) suggested a fuzzy-based technique for image inpainting that eliminated the dropping effect associated with exemplar-based inpainting [17].

An algorithm that disperses the median value of pixels from the outside region into the painted area was proposed by Thanh et al. (2019) [18]. It is a reliable technique with encouraging outcomes for both homogeneous and heterogeneous backgrounds.

Wang et al. (2006) suggested a technique for image inpainting that depends upon Compactly Supported Radial Basis Function (CSRBF) [19]. This algorithm translated 2D picture inpaint into an implicit surface reconstruction problem using a 3D point set. The RBF algorithm reduces the computational cost of the sparse and bounded linear algebraic equation system. Zhang N et al. (2019) presented an inpainting algorithm based on the exemplar technique [20]. Sun et al. (2005) offer an interactive curve-based strategy to complete essential structures before remaining undetermined parts [21]. A new randomized method called PatchMatch is presented by Barnes et al. (2009) for rapidly determining the estimated nearest neighbour or any matches found in the patch images [22]. These methods work well, especially when used for inpainting stationary backgrounds with repeating patterns. However, these methods might not be able to fill up large gaps in complex scenes.

Original image is separated into textured and non-textured sections by Hung et al. (2017) encompassing the damaged area components using a structural tensor [23]. Exemplar-based restoration is used for textured regions, while Telea restoration is used for non-textured sections [7]. Specific Cubic Spline interpolation technique is used to fill curve sections [24]. This method's disadvantage is that using so many algorithms increased the computing load. When resampling images, the nearest-neighbour, linear, and different cubic interpolation functions are commonly utilized. For the most part, quadratic functions have been ignored since it is believed that they create phase distortions [25].

In traditional image inpainting method, image inpainting techniques fill the missing portion primarily using statistical

data from the remaining image content. To maintain consistency with the surrounding pixels, each pixel of the missing portion is constructed using the similarity principle. However, when the data dimension increases, interpolation-based techniques, which can solve the problem with great precision, may lose their effectiveness [26].

These classical methods like image interpolation, exemplar-based, and PDE-based techniques cannot recreate complicated or heavily corrupted areas [27].

Image inpainting based on non-learning

Non-learning-based inpainting approaches have drawbacks in comparison to deep learning-based techniques when it comes to handle complicated textures, semantically meaningful inpainting, and dealing with vast missing sections. They can be computationally efficient and may be preferred only in cases where training data is limited or when a deterministic approach is needed. Significant progress has been made for creation of supervised learning methods for image inpainting with deep learning. A specific kind of machine learning called deep learning is primarily concerned with teaching computers through experimentation. Deep image inpainting models can produce more convincing material for complex situations than non-learning-based alternatives [27].

Convolutional neural network will be trained to detect missing pixels in a damaged image by training the machine for achieving effective image inpainting. By utilizing data from distant image contexts, deep image inpainting models deduce the contents of a large missing region. An artificial convolutional neural network trained to produce contents of missed area based from environment [28]. Early implementations of the Context Encoder concept have demonstrated encouraging outcomes using images of people's faces, streets, etc. These models are limited to processing low-resolution images since they employ fully-connected layers. Since this concept is introduced, we will refer to it as partial convolution-based padding [29]. It lets us perceive the padded region as holes while keeping the original image as available area. Considering the ratios between the sliding window area's convolution and padded areas, convolution technique properly reweights convolution results around image borders [30,31]. In this paper, the missing material is generated by conditioning on the available data which use a unique method for semantic image inpainting [32]. They use the context and past losses, along with a trained generative model to repair damaged image. The generative model is then applied to this encoding to infer the content that is missing. While the cutting-edge learning-based method requires precise knowledge about the gaps in the training phase, the method allows inference regardless of how the missing content is constructed. Context encoders that have been taught to produce context-based images advance the field of conceptual inpainting. Simultaneously, they acquire feature

representations that can compete with those of other models trained with additional guidance [33]. The development of adversarial training and deep feature learning for picture inpainting has resulted in notable advancements. Deep image inpainting models are more capable of producing more credible contents for complex situations as compared to non-learning based systems [34]. MagConv presents a unique convolution method specifically suited for image inpainting [14]. This method contains learnable piecewise activation function and sharing of the convolution kernel between mask and image. Although this method gives plausible results, this method contains more complicated and numerous instructions to compute. In paper is made up of an adversarial model-based image completion network and an edge generator [35]. When a sizable portion of an image is absent or a lot of texture is present in image, especially in higher quality photographs, this edge generating model occasionally has trouble accurately representing edges [36]. Reference used PConv to represent the combined processes of texture-guided structure reconstruction and structure-constrained texture generation [13]. To further enhance PConv's performance, alternate methods such as Gated Convolution (GConv) and Learnable Bidirectional Attention Maps (LBAM) were proposed. A stimulating attention map module for mask replacement is presented by LBAM with a feature re-normalization. A soft multi-channel mask is learned by GConv in order to re-normalize features. Unlike PConv, which scales only the hole border, these 2 algorithms scale the features of the entire image. Additionally, region-wise solutions were put out to use various convolution kernels in the decoder network to learn distinct properties of valid regions and holes independently [37-40].

From the literature review, we reached in a conclusion that these non-learning based image inpainitng methods fail in understanding the content in the image. Many non-learning based inpainting techniques require manual tuning of parameters such as patch size or diffusion coefficients. This procedure could take a long time and produce poor results. Moreover, these traditional methods are mainly focused on size of the region and also these methods find difficulty in handling irregular inpainting regions.

When compared to conventional techniques, learning based algorithms acquire and recognize advanced semantic characteristics of images in order to create efficient image inpainting, which is required for scenes with plenty of gaps and complexity.

Methodology

As depicted in Figure 1, our suggested approach comprises two important operations; one as a convolution operation and the other one as a max-pool operation. In this portion, we explain about our suggested method in detail. Afterwards, network architecture along with loss function associated with the work are discussed.
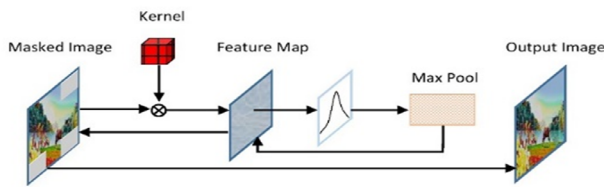
Figure 1: Architecture of the suggested method, ReConv

Image inpainting using repeated convolution process (ReConv)

In recent existing methods, they all treat valid elements (valid elements in an image refer to the parts of the image that contain meaningful information) and invalid elements (refer to the parts of the image that are missing portions) of the image separately [13,14]. This approach makes the system more complex in terms of space and time. In our proposed method, we do a repeated convolution operation by treating valid and invalid elements alike which reduces the complexity of the system. Convolution is a technique of altering an image by running a kernel matrix sliding upon each and every pixel of the input image.

Currently, there are no available dataset of masked images as online, but in the case of image inpainting a dataset of masked images is necessary for testing the proposed model, therefore, we created Masked images (M) of size 'n' for inpainting.

Mask creation for images involve marking the areas of an image that we want to correct or fill in. In this method, mask of an image is produced by taking copy from the original image and make background image as white. Then, select random (x, y) points in the image and draw lines by joining these (x, y) points in random thickness of different orientations with customized number of lines.

Resultant feature map matrix after convolution is subjected to the max pooling process. By producing a down sampled (pooled) feature map, the resultant max-pooled matrix determines the largest value of each patch in the feature map. Consequently, the max-pooling layer's feature map would include the standout elements of the initial feature map obtained using equation (1).

Then size of resultant down sampled matrix (Mx × Mx) after max pooling operation has the size n^2/l^2 where 'n' size of 'F' and 'l' size of 'k'.

$$mx = \frac{n^2}{l^2} \qquad \text{---------------- (1)}$$

or

$$mx = \frac{floor(n-f)}{s} + 1 \qquad \text{------- (2)}$$

'n' represents the size of masked image and 'l' is the max pool's filter size operation and 's' is the stride and Mx is the size of the max pooled matrix [41]. For each element in the Max-pooled matrix (Mx) locate the matching place in feature map matrix and associated data element in the masked image to determine each element's location in the masked image. After finding a data element in the masked image, search for a nearby zero in the masked image. Then replace this zero element with this data element. After one

replacement check the updated masked image for invalid element (zero). If there is invalid element (zero) perform the above operations again and again until no invalid (zero) in the updated masked image, then the updated masked image is the new inpainted image.

Auto encoder network architecture and implementation

As displayed in Figure 2 we design a U-Net like architecture including a decoder and an encoder. The damaged image must be encoded and transformed into latent feature maps by the encoder, and the decoder must extract the image from these latent feature forms. We stacked 8 layers of convolutions in the encoder part and 8 layers of convolutions at the decoder part. The following convolution layer will get the masked picture as input from the skip link, which will concatenate the mask with the original image. The final convolution layer enables the model to replicate non-hole pixels from the regions of the image that are accessible. The section on detailed configuration is explained in Table 1. We used Adam optimizer for optimization purpose. We train a batch size of 32 on a single *NVIDIA GIV (16GB) [42,43].

Table 1: Layered architecture of suggested model reconv

Layer	Output	No. of channels	Parameter	Activation Function
Input Image	128 × 128	3	0	Relu
Encoder	128 × 128	3	0	Relu
Conv1	128 × 128	3	1760	Relu
TFOpLambda	128 × 128	32	0	Relu
Conv2	64 × 64	32	18464	Relu
TFopLambda	64 × 64	32	0	Relu
Conv3	64 × 64	32	18464	Relu
TFopLambda	64 × 64	32	0	Relu
Conv4	32 × 32	32	18464	"
TFOpLambda	32 × 32	32	0	"
Conv5	32 × 32	32	18464	"
Conv6	16 × 16	32	18464	"
TFOpLambda	16 × 16	32	0	"
Conv7	16 × 16	32	18464	"
TFOpLambda	16 × 16	32	0	"
Max-pooling2D	3 × 3	3	0	"
Conv2D	8 × 8	32	0	Relu
TFOpLambda	8 × 8	32	0	Relu
Upsampling 2D	16 × 16	32	0	Relu
Upsampling 2D	16 × 16	32	0	Relu
Concatenate	16 × 16	64	0	Relu
Concatenate	16 × 16	64	0	Relu
Conv2D	16 × 16	256	295168	Relu
TFOpLambda	16 × 16	256	0	Relu
Conv2D	16 × 16	128	589952	Relu
TFOpLambda	16 × 16	128	0	Relu

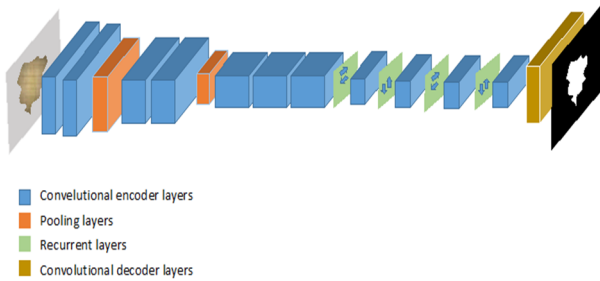|				
Upsampling 2D	32 × 32	128	0	Relu
Upsampling2D	32 × 32	128	0	Relu
Concatenate	64 × 64	96	0	Relu
Concatenate	64 × 64	96	0	Relu
Conv2D	64 × 64	32	36896	"
TFopLambda	64 × 64	64	0	"
Conv2D	64 × 64	32	0	"
TFopLambda	128 × 128	3	0	"
Conv2D	128 × 128	3	84	"
Total parameters	17,18,679	-	-	-
Total trainable parameters	17,18,679	-	-	-

Figure 2: Basic Network diagram of Auto Encoder

Loss function

Our loss functions aim to achieve 2 goals: Composition (i.e., how well expected missed pixel values meld with the surrounding area) and per-pixel reconstruction accuracy. The total of all absolute deviations between the value that is true and the value that is expected is the error. This disparity is minimized by applying L1 loss function. Mean of these Absolute Errors (MAE), often known as the Mean Absolute Error or L1 loss is employed to convey the loss function [44].

The goal is to reduce this loss between expected and desired outputs as much as possible throughout training. MAE results by dividing the total absolute errors by the sample size.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad \text{---------- ------- (3)}$$

To replicate how people might see certain aspects of an image and to record high-level semantic features, the perceptual loss is described as [45].

$$L_{perc} = \sum_{i=1}^{N} \| \Psi_i(I_0) - \Psi_{gt}(I_{gt}) + \sum_{i=1}^{N} \| \Psi_i(I_{comp}) - \Psi_{gt}(I_{gt}) \cdot (4)$$

where Icomp=M*k representing the convolutional process over the masked image and '*' is the convolution symbol. Ψi stands for the feature map of the ith pooling layer. Once the loss functions mentioned above are gathered, the whole loss function may be expressed as,

$$Total\ loss = MAE + L_{perc} \quad \text{----------} \quad (5)$$

ReLU is the activation function that's being employed here for reducing loss. ReLU, sometimes referred to as the rectified linear activation function that supports non linearity. In the case of a positive input, ReLU will output the input directly, in the absence of a positive input, it will result into zero. So it reduces complexity. It has taken as a standard activation function for numerous types of neural networks since models that utilize it typically perform better and train more quickly [46].

The details of image inpainting datasets and evaluation metrics are covered in this section. We also made a systematic comparison and the performance of our suggested image inpainting model in comparison to inpainting techniques in literature.

Experimental Results

Data sets

We created irregular masks for image inpainting by adding random lines to the training data. We selected irregular masks because irregular masks are challenging and more applicable to real-life circumstances mainly used by majority of the image inpainting techniques. We experimented on 2 datasets with varying thickness of the mask. Two popular datasets used are CelebA and CIFAR-10 and these are employed for training separately, and the prepared model was then used for testing. It is evident that our approach can faithfully recreate the features, including its shape and texture.

CelebA: We employed CelebA (Large-scale CelebFaces Attributes) 128 × 128 dataset, a sizable collection of 30,000 images. 80% of total images are taken for training and 20% of the total images are taken for testing [47].

CIFAR-10: CIFAR-10 dataset contains variety collection of images specially designed for machine learning and computer vision algorithms. We use CIFAR-10 image dataset made up of 60,000 images with size 32 x 32 photos with 50,000 images for training and 10,000 samples for testing purpose [48].

We first activate initial training with a learning rate of 0.0002 before using batch normalization when holes are present. Next, after freezing the batch normalization parameters in the encoder portion of the network, we use a learning rate of 0.00005 to fine-tune them. By maintaining batch normalization enabled on the decoder part helps to expedite the convergence process and prevent issues with the wrong mean and variance.

Experimental setup

The proposed image inpainting model was trained on Google Colab pro platform with 54.8 GB RAM with Tesla v100 GPU machine with Keras libraries. The Adam optimizer, sometimes referred as "Adaptive Moment Estimation", is an iterative optimization method used in neural network training that reduces the loss function [43]. It was utilized to train the model across 40 epochs.

Evaluation metrics

The Structural Similarity Index Measure (SSIM) and FID (Frechet Inception Distance) metrics, as well as the Peak Signal to Noise Ratio (PSNR), are the frequently employed to assess the strength of images. Below is a list of all the objective measurements that were utilized in the quantitative comparisons along with explanation of their selection [49].

Peak Signal to Noise Ratio (PSNR): PSNR between 2 images is an expression for ratio of signal power to noise power. Using this ratio, the quality of original image and inpainted images is compared. As PSNR rises, the restored image's quality gets better.

$$PSNR(p,q) = 10 \log_{10} \frac{255^2}{MSE(p,q)} \quad ------ \quad (6)$$
$$MSE = \sum(y_i - p_i)^2 \times n \quad ------ \quad (7)$$

in which 'n' is the quantity of r of data points, pi is the observed value, and yi is the forecasted value.

Structural Similarity Index Measure (SSIM): SSIM is utilized as one of the most representative quality measures in many fields of image processing. When the SSIM value is nearly equal to '1' it indicates better structural similarity between inpainted and original image.

$$SSIM(i,j) = \frac{2(\mu_i \, \mu_j + c_1)(2c_{ij} + c_2)}{(\mu_i^2 + \mu_j^2 + c_1)(a_i^2 + a_j^2 + c_2)} \quad ----- \quad (8)$$
$$c_1 = (k_1 L)^2 \qquad c_2 = (k_2 L)^2$$

c1, c2-variables to stabilize the division with weak denominator

L is the dynamic range of pixel values (typically this is $2^{\#bits \, per \, pixel} - 1$, k_1=0.01 k_2=0.03 by default [49].

FID: A metric called Frechet Inception Distance score (FID) determines the separation between feature vectors computed for generated and real images.

$$FID = (mu1 - mu2)^2 + T_r(c_1 + c_2 - 2*sqrt(c_1*c_2) - \quad (9)$$

Where mu1 and mu2 stand for the original and produced images' feature-wise means. The trace linear algebra operation is denoted by Tr, and the co-variance matrix for original and produced feature vectors is represented by c1

and c2 [50].

Quantitative comparisons

Tables 2 and 3 shows the PSNR, SSIM and FID values of increasing order of epochs for our proposed model. Tables 4 and 5 shows these values of different mask ratios. Quantitative analysis of image inpainting in deep learning evaluates the effectiveness of inpainting models by utilizing a variety of matrices. These tables depict by comparing with state-of-art methods EC, CTSDG and MGConv on CelebA dataset [14,51,52]. From these tables, it reveals that the suggested method shows notable advancement in PSNR, SSIM and FID values, which implies excellence of our method for inpainting irregular large holes. Moreover, we have given values for these matrices for different mask ratios. In every mask ratio, recommended method shows superior outcomes in comparison to the existing methods in literature.

Table 2: PSNR AND SSIM values of different values of epoch on CIFAR-10 DATASET

Epoch	20	25	30	35	40
PSNR	20.673	20.435	21.234	21.712	25.101
SSIM	0.926	0.927	0.938	0.937	0.947
FID	2.022	1.98	1.99	1.88	1.834

Table 3: PSNR AND SSIM values of different values of epoch on CELEBA DATASET

Epoch	20	25	30	35	40
PSNR	23.872	23.011	24.213	24.658	25.101
SSIM	0.956	0.967	0.978	0.978	0.988
FID	1.912	2.021	1.956	1.823	1.612

Table 4: PSNR, SSIM and FID values of the proposed method recent quantitative measures are not available for CIFAR-10 DATASET for comparison

-	Mask ratio	ReConv (our method)
SSIM↑	0%-20%	0.9513
	20%-40%	0.9532
	40%-60%	0.9223
PSNR↑	0%-20%	34.236
	20%-40%	28.962
	40%-60%	21.101
↓FID	0%-20%	2.452
	20%-40%	6.125
	40%-60%	11.831

Table 5: PSNR,SSIM AND FID values with state of the art in quantitative comparison with celeba dataset, ↑ suggests that greater is preferable and ↓suggests lower is preferable

| -	Mask ratio	EC	CTSDG	MGConv	ReConv (proposed method)
SSIM↑	0%-20%	0.9908	0.9908	0.9904	0.9913
	20%-40%	0.953	0.9572	0.9576	0.9592
	40%-60%	0.8633	0.8747	0.8828	0.8923
PSNR↑	0%-20%	36.752	38.17	37.67	38.236
	20%-40%	28.851	29.48	29.299	29.962
	40%-60%	23.67	24.205	24.268	25.101
↓FID	0%-20%	2.374	2.29	1.815	1.712
	20%-40%	6.707	8.209	5.616	5.125
	40%-60%	13.151	17.519	10.554	9.231

Qualitative comparisons

On 2 datasets, Figures 3-7 provide a visual comparison of our strategy with some samples taken from the result. The true image is shown in the first column, while masked image and mask are shown for each dataset in the following 2 columns and finally the last column shows the final inpainted image by our proposed model. While most of the algorithms are able to effectively rebuild the bulk of the missing pixels in the gaps, some patterns are distorted and some images have intensity contrast mismatches. By looking into the visuals given, it can be derived that our method, repeated convolution with max-pooling (ReConv) works well and generates pleasing retrieval of structure and texture of images [53].

Learning curves

Figure 3 displays learning curves obtained at the time of training using CIFAR-10 dataset and CelebA dataset. It shows 'Mean Absolute Error' loss function which clearly lowers with an increase in training epochs. It also indicates that the network may learn more until it reaches the convergence state. At the beginning of training, the loss is considerable, but by the end, the loss decreases after 20th epoch and settles within the 40th epoch range.
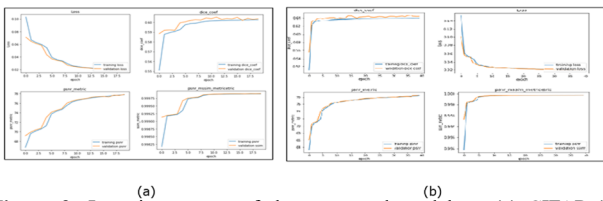
Figure 3: Learning curves of the suggested model on (a) CIFAR-10 dataset and (b) CelebA dataset
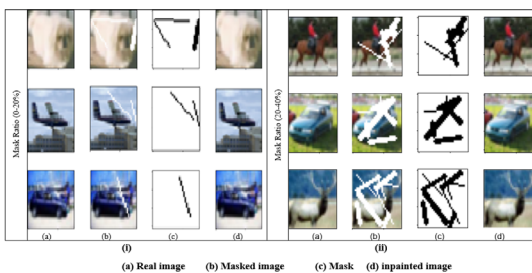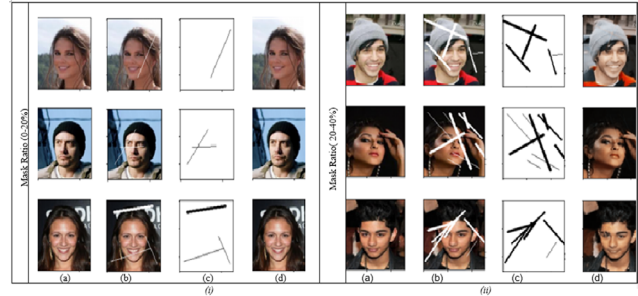

Figure 4: (i) and (ii) represents images having mask ratios 0%-20% and 20%-40%; (a) represents the image that appears originally and (b) represents image by applying mask on it, (c) is the visualization of how mask appears, (d) is the resultant image after applying the ReConv method on CIFAR-10 data set with increased thickness of mask
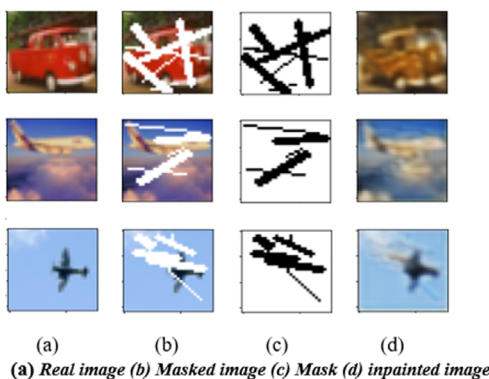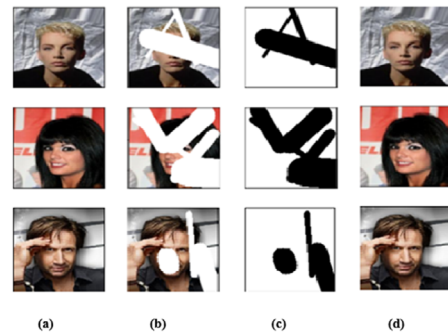

Figure 5: Mask ratio having 40%-60%; (a) represents the image that appears originally and (b) represents image by applying mask on it, (c) is the visualization of how mask appears, (d) is the resultant image after applying the ReConv method on CIFAR-10 data set with increased thickness of mask

Figure 6: (i) represents images having mask ratios 0%-20%, (ii) represents images having mask ratios 20%-40%; (a) represents the image that appears originally and (b) represents image by applying mask on it, (c) is the visualization of how mask appears, (d) is the resultant image after applying the ReConv method on CelebA data set with increased thickness of mask

Figure 7: Images having mask ratios (40%-60%); (a) represents the image that appears originally and (b) represents image by applying mask on it, (c) is the visualization of how mask appears, (d) is the resultant image after applying the ReConv method on CelebA data set with increased thickness of mask

Ablation study

Here, we provided a novel method for efficiently inpainting images using repeated convolution method with max-pooling, named ReConv, which effectively inpaint images. Our model is capable of handling holes that are any size, shape, location, or distance from the edges of the image. We have experienced the model for different values of epochs. From the experiment, we identified that in 40th epoch, the model is converging. Peak signal-to-noise ratio: A decibel measurement, is computed by PSNR block in between 2 images. The increased value of this ratio is used to evaluate the effectiveness of our approach. As the PSNR rises, the rebuilt image's quality gets better. In addition to this, we calculated SSIM and FID measures which shows the practicability of our method.

Discussion

Inpainting is useful in reconstructing 3D models from partial data, such as missing sections in MRI or CT scans, ensuring more complete anatomical models for diagnosis and surgical planning. Our method does not contain any complicated computations which only uses a standard convolution operation with max-pooling. We use irregular mask of random thickness, in any orientation and number

of lines also varying.

Conclusion

The scope of image inpainting in drug and alcohol addiction research is wide-ranging, encompassing clinical therapy, medical imaging, research ethics, and public awareness. By improving therapeutic interventions, ensuring privacy, and augmenting medical analysis, inpainting offers significant potential for both treatment providers and researchers, advancing the understanding and management of addiction. Inpainting can be used to alter or mask personal information (such as facial features) in medical imaging data while preserving critical medical content. This enables compliant data sharing for research purposes while ensuring patient privacy. In the current investigation, we suggested a new approach for effective image inpainting by using repeated convolution method. Our model can manage holes that are any size, shape, location, or distance from the edges of the image. Furthermore, we tested images of different type of random masks of varying thickness and length. The experimental outcomes on CelebA and CIFAR-10 dataset demonstrated the viability of our approach. We showed how repeated convolution with max-pooling can enhance the level of image inpainting quality. We compared our model's efficacy to alternative image inpainting techniques. Our model produces outcomes with rich texture and consistent structure when compared to existing approaches.

Acknowledgement

None.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, Image inpainting. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques, (2000):417–424.

2. A.K. Al-Jaberi, E.M. Hameed, A review of PDE based local inpainting methods, (2021):012149

3. M. Bertalmio, L. Vese, G. Sapiro, S. Osher, Simultaneous structure and texture image inpainting, IEEE Trans Image Process, 12(2003):882-889.

4. A. Criminisi, P. Perez, K. Toyama, Region filling and object removal by exemplar-based image inpainting, IEEE Trans Image Process, 13(2004):1200-1212.

5. L.J. Deng, T.Z. Huang, X.L. Zhao, Exemplar-based image inpainting using a modified priority definition, PLOS One, 10(2015):0141199.

6. M. Shroff, M.S.R. Bombaywala, A qualitative study of exemplar based image inpainting. SN Applied Sciences, 1(2019):1-8.

7. A. Telea, An image inpainting technique based on the fast marching method, J Graph Tools, 9(2004):23-34.

8. I.N. Sari, E. Horikawa, W. Du, Interactive image inpainting of large-scale missing region, IEEE Access, 9(2021):56430-56442.

9. L. Chang, Y. Chongxiu, New interpolation algorithm for image inpainting, Phys Procedia, 22(2011):107-111.

10. P. Jini, P, K.K. Rajkumar, Image inpainting using image interpolation-an analysis, Revista Geintec, 11(2021):1906-1920.

11. M. Zor, E. Bostanci, M.S. Güzel, E. Karataş, analysis of interpolation-based image in-painting approaches, in advanced sensing in image processing and IOT, (2022):153-170.

12. Y. Zeng, J. Fu, H. Chao, B. Guo, Aggregated contextual transformations for high-resolution image inpainting, IEEE Trans Image Process, 29(2022): 3266-3280.

13. G. Liu, F.A. Reda, K.J. Shih, T.C. Wang, Image inpainting for irregular holes using partial convolutions, Comput Vis ECCV, (2018):85-100.

14. X. Yu, L. Xu, J. Li, X. Ji, MagConv: Mask-guided convolution for image inpainting, IEEE Trans Image Proc, 32(2023):4716-4727.

15. U. Sara, M. Akter, M.S. Uddin, Image quality assessment through FSIM, SSIM, MSE and PSNR-a comparative study, J Comput Commun, 7(2019):8-18.

16. Z. Wang, K. Qinami, I.C. Karakozis, K. Genova, P. Nair, Towards fairness in visual recognition: Effective strategies for bias mitigation, (2020):8919-8928.

17. M. Ghayoumi, C.C. Lu, Improving exemplar based inpainting method with a fuzzy approach. In 2014 International conference on audio, language and image processing, IEEE, 19(2014):671-675.

18. D.N. Thanh, V.S. Prasath, H. Kawanaka, Image inpainting method based on mixed median, IEEE, 7(2019):24-29.

19. W. Wang, X. Qin, An image inpainting algorithm based on CSRBF interpolation, Int J Inf Tech, 12(2006):112-119.

20. N. Zhang, H. Ji, L. Liu, G. Wang, Exemplar-based image inpainting using angle-aware patch matching, Eurasip J Image Video Process, 70(2019),1-13.

21. J. Sun, L. Yuan, J. Jia, H.Y. Shum, Image completion with structure propagation, ACM SIGGRAPH, (2005):861-868.

22. C. Barnes, E. Shechtman, A. Finkelstein, PatchMatch: A randomized correspondence algorithm for structural image editing, ACM Trans Graph, 28(2009): 619-629.

23. N.V. Hung, N.T.T. Hien, P.T. Vinh, N.T. Thao, An utilization of edge detection in a modified bicubic interpolation used for frame enhancement in a camera-based traffic monitoring, Int Conf on Inf Comm, 7(2017):316-319.

24. M. Motmaen, M. Mohrekesh, M. Akbari, N. Karimi, Image inpainting by hyperbolic selection of pixels for two-dimensional bicubic interpolations, Elec Eng, 17(2018):665-669.

25. B. Vadhel, B. Limbasiya, Survey on different techniques for image inpainting, Int Res J Tec, 3(2016): 852–882.

26. H. Xiang, Q. Zou, MA. Nawaz, X. Huang, F. Zhang, et al. Deep learning for image inpainting: A survey, Pat Rec, 134(2023):109046.

27. J. Jam, C. Kendrick, K. Walker, V. Drouard, J.G.S. Hsu, et al. A comprehensive review of past and present image inpainting methods, Comp Vis Im Und, 203(2021):103147.

28. K. Oshea, R. Nash, An introduction to convolutional neural networks, ar pre ar, 1511(2015):08458.

29. D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, AA. Efros, Context encoders: Feature learning by inpainting, (2016):2536–2544.

30. G. Liu, KJ. Shih, TC. Wang, FA. Reda, K. Sapra, et al. Partial convolution based padding, 1811(2018):11718.

31. G. Liu, A. Dundar, KJ. Shih, TC. Wang, FA. Reda, et al. Partial convolution for padding, inpainting, and image synthesis, 45(2022):6096-6110.

32. AVD. Oor, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al. Conditional image generation with pixelcnn decoders, (2016):4790–4798.

33. Y. Zeng, J. Fu, H. Chao, B. Guo, Learning pyramid-context encoder network for high-quality image inpainting, (2019):1486-1494.

34. T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, et al. Improved techniques for training gans, (2016):2234–2242.

35. V. Chandak, P. Saxena, M. Pattanaik, G. Kaushal, Semantic image completion and enhancement using deep learning, ICCCNT, 10(2019):1-6.

36. D. Wang, C. Xie, S. Liu, Z. Niu, W. Zuo, Image inpainting with edge-guided learnable bidirectional attention maps, Ar Pre Ar, 10(2021):1026–1034.

37. J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, 25(2012).

38. P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, et al. Understanding convolution for semantic segmentation, WACV, (2018):1451-1460.

39. S. Park, Y.G. Shin, Generative convolution layer for image generation, Neu Ne, 152(2022): 370-379.

40. A.Y. Raymond, C. Chen, T.Y. Lim, GS. Alexander, M.H. Johnson, et al. Semantic image inpainting with perceptual and contextual losses, Ar Pre Ar, 3(2016):1607.07539.

41. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition, 40(2017):1452–1464.

42. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, Spr Int Pub, 9351(2015):234-241.

43. Z. Zhang, Improved adam optimizer for deep neural networks, IWQoS, (2018):1-2.

44. T.O. Hodson, Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): When to use them or not, Ge Mod De Dis, 15 (2022):1-10.

45. Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss, 37(2018):1348-1357.

46. A.F. Agarap, Deep learning using rectified linear units, Relu, 2(2019):1803.

47. Z. Liu, P. Luo, X. Wang, X. Tang, Large-scale celebfaces attributes (celeba) dataset, 15(2018):11.

48. A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM. In 2010 20th International conference on pattern recognition, (2010): 2366-2369.

49. Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, Conf Comput Vis, (2015): 3730–3738.

50. Z. Shangguan, Y. Zhao, W. Fan, Z. Cao, Dog image generation using deep convolutional generative adversarial networks, (2020): 24-27.

51. K. Nazeri, E. Ng, T. Joseph, F. Qureshi, M. Ebrahimi, Edgeconnect: Structure guided image inpainting using edge prediction, 2019.

52. X. Guo, H. Yang, D. Huang, Image inpainting *via* conditional texture and structure dual generation, 2021: 14134-14143.

53. W. Huang, Y. Deng, S. Hui, J. Wang, Image inpainting with bilateral convolution, Rem Se, 14(2023): 6140.