Rejuvenate: Image Inpainting

Aaditya Baranwal
B20EE001
baranwal.1@iitj.ac.in

Vishal Kumar Yadav
B20EE080
yadav.40@iitj.ac.in

Abstract

This paper provides an overview of image inpainting, an important technique in the field of computer vision and image processing. The paper discusses the significance of image inpainting, various methods and its applications. It provides an overview of various techniques used for image inpainting, partial differential equation-based methods, as well as deep learning-based methods. The paper then reviews two classical methods for image inpainting, the Partial Differential Equation Inpainting method based on Image Characteristics and the Harmonic Inpainting method. The paper also reviews two deep learning bases methods - MAT- Large Hole Image Inpainting and Context Encoder GAN.

GitHub:*https://github.com/eternal-f1ame/Rejuvenate*

1. Introduction

Image inpainting is a widely studied and important technique in the field of computer vision and image processing. Its objective is to recover missing or damaged regions of an image in a visually plausible and consistent manner. This technique has various applications, ranging from restoring old photographs to removing unwanted objects from images. In this article, we will discuss the significance of image inpainting, its applications, challenges faced by researchers in this field, and the different techniques used for image inpainting.

The need for image inpainting arises when images become corrupted or damaged due to various reasons, such as sensor noise, compression artifacts, or occlusions caused by objects or people. In such cases, it becomes challenging to interpret the visual content of the image without the missing regions. Image inpainting algorithms aim to fill in these gaps, enabling better visual interpretation of the image and enabling further processing or analysis.

Image inpainting has numerous applications, ranging from restoring old photographs to removing unwanted objects from images. It has been used in the film industry to remove wires or rigs used during filming. In the medical field, image inpainting has been used to remove artifacts from medical images caused by patient motion or equipment malfunctions. Additionally, image inpainting has also been used in forensic science to restore blurred or distorted images.

Over the years, several techniques have been proposed for image inpainting, including traditional methods based on texture synthesis, patch-based methods, partial differential equation-based methods, and more recently, deep learning-based methods. Traditional techniques often rely on hand-crafted features and are limited by their inability to capture the complex structure and context of natural images. In contrast, deep learning-based methods leverage the power of deep neural networks to learn the underlying patterns and structures of images, resulting in more realistic and visually plausible inpainting results.

One of the challenges faced by researchers working on image inpainting is the unavailability of a large dataset of damaged painting images. This presents a significant obstacle to the development of image inpainting methods, as these methods require large amounts of labeled training data to learn the underlying patterns and structures of images. However, due to the novelty of the image inpainting topic, such datasets are not readily available.

In this paper we reviewed different methods for image inpainting, both classical and Deep Learning Based. We reviewed the Partial Differential Equation Inpainting method based on Image Characteristics that takes the gray value of each pixel as heat, and take the image inpainting processing as heat diffusion. Next we review Harmonic Inpainting that extends the harmonic function from the boundary of the missing domain into the interior to fill in missing parts of the image.

2. Related Work

2.1. Classical Image Inpainting Methods:

Texture Synthesis-based methods, patch-based methods. Texture synthesis-based methods fill in missing regions by synthesizing the texture from the surrounding regions. Patch-based methods use patches from the surrounding regions to fill in the missing regions. PDE-based methods- [4] solve a diffusion process to recover the missing regions.

Harmonic inpainting- [3] is a popular PDE-based method that uses the Laplacian operator to restore damaged or missing regions in the image. Other methods include Interpolation methods like nearest-neighbor, bilinear, and bicubic interpolation, Statistical Methods like maximum a posteriori (MAP) estimation and Bayesian image restoration, etc.

2.2. Deep Learning-based Image Inpainting Methods:

Deep learning-based methods leverage the power of deep neural networks to learn the underlying patterns and structures of images. Generative Adversarial Networks (GANs), Autoencoders, and Convolutional Neural Networks (CNNs) are some of the popular deep learning-based image inpainting techniques. GANs have shown promising results by generating realistic and visually plausible inpainting results. Autoencoders use an encoder-decoder architecture to learn the underlying patterns of images and generate the missing regions. CNN-based methods use convolutional layers to learn the features of the images and generate the missing regions.

2.2.1 GAN based Image inpainting

Generative Adversarial Networks (GANs) have been successfully used for image inpainting tasks due to their ability to generate high-quality, visually plausible results. GAN-based methods for image inpainting can be broadly categorized into two types: the traditional GAN-based methods and the Conditional GAN (cGAN)-based methods.

Traditional GAN-based methods for image inpainting involve training a generator network to generate the missing regions of an image given the input image and the mask of the missing regions. The discriminator network is trained to distinguish between the generated image and the ground truth image.

Conditional GAN-based methods for image inpainting overcome the limitations of traditional GAN-based methods by conditioning the generator and discriminator networks on the input image and the mask of the missing regions. The generator network takes both the input image and the mask as input and generates the missing regions of the image. The discriminator network takes both the input image and the generated image as input and is trained to distinguish between the generated image and the ground truth image. Conditional GAN-based methods have been performant and have shown to be able to preserve the fine details of the image, even when multiple missing regions are present.

One of the most popular GAN-based methods for image inpainting is the Context Encoder (CE) method proposed by Pathak et al. in 2016 - [2]. The CE method uses a GAN to generate the missing regions of an image by conditioning the generator and discriminator networks on the input image and the mask of the missing regions. The CE method has been shown to generate high-quality results and preserve the fine details of the image, even when multiple missing regions are present.

2.2.2 Transformer-based Image Inpainting Methods

Recently, transformer-based methods have been applied to image inpainting tasks. The transformer architecture is a neural network that is composed of a series of self-attention layers. These layers can capture long-range dependencies. The self-attention mechanism in transformers allows the model to weigh the importance of different regions of the image when generating the missing regions. This enables the model to generate visually plausible and consistent results.

The Mask-Aware-Transformer is a popular transformer-based image inpainting method that uses the self-attention mechanism to generate the missing regions. It uses the input mask to generate the attention maps, which capture the contextual information of the image. The attention maps are used to generate the missing regions. Recent advancements: MAT of Large Hole Image Inpainting - [1] has set up new benchmarks.

An advantage of transformer-based methods is their ability to handle arbitrary shapes and sizes of missing regions. Traditional methods often rely on patch-based approaches, which may not be able to handle irregular shapes or sizes of missing regions.

One popular transformer-based image inpainting method is the Mask-Aware-Transformer. This method uses the input mask to generate the attention maps, which capture the contextual information of the image. The attention maps are then used to generate the missing regions.

3. Methodology and Implementations

3.1. PDE inpainting based on image Characteristic

PDE image inpainting method considers the gray distribution of image as heat distribution and regards the process of image evolution as heat conduction. By establishing and solving the corresponding PDE model to accomplish heat diffusion, image gray will be redistributed . PDE image inpainting method interpolates the missing gray values for the initial image through gray diffusion.

In Fig. 1, assuming Ω as the entire image domain of the image I, D is the inpainting domain, E is the neighbourhood out of the inpainting domain, $\mathcal{T}$ is the boundary of D.

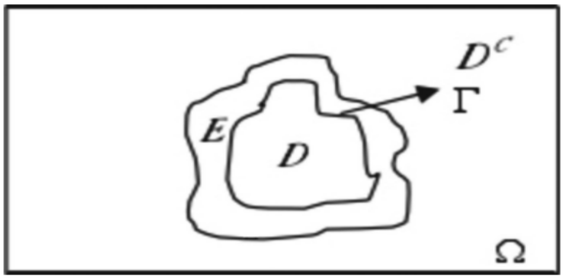

Figure 1. Inpainting Domain and neighbourhood

Let u denote the repaired image.
The following PDE is used for the image inpainting process-

$$\partial_t u = g(C_u)(g(|\nabla u|)D^2(\eta,\eta) + \alpha D^2(\mathcal{E},\mathcal{E})), (x,y) \in D$$
$$u = I \quad (x,y) \in D^C$$

where $u(x,y,t)$ is the evolving image, whose initial values is $u(x,y,0) = I(x,y)$, ∇u is the gradient of u, C_u is the corner intensity of u; $g(x) = 1/(1+x^2/k^2)$, thus $g(|\nabla u|)$ and $g(C_u)$ are the edge stopping function and the corner protection operator respectively; $D^2(\eta,\eta)$ and $D^2(\mathcal{E},\mathcal{E})$ are second-order directional derivatives in the gradient direction $\eta = \nabla u/|\nabla u|$ and tangential direction $\mathcal{E} = \nabla^{\perp} u/|\nabla u|$. We update the image as:

$$u^{n+1} = u^n + \Delta t.\partial_n u$$

3.2. Harmonic Image Inpainting

Harmonic inpainting is a type of partial differential equation (PDE) method for image inpainting. It involves extending the harmonic function from the boundary of the missing domain into the interior to fill in missing parts of an image. The inpainted image is computed as a solution of the Laplace equation or as a minimiser of the Dirichlet energy over the inpainting domain. It constitutes a smooth and linear interpolation process that roughly fills in missing grey values by averaging the given grey values on the boundary of the inpainting domain.
Let $g \in L^2(\Omega)$ be the given image, whose grey values got lost inside the inpainting region $D \subset \Omega$. The inpainted image is computed as a weak soluton of the Laplace Equation

$$\Delta u = 0,\ in\ D,$$

$$u = g,\ on\ \partial D$$

that is, u fulfils

$$\int_D \nabla u \nabla \psi dx = 0,\ \forall \psi \in H_0^1(D)$$

,

3.3. MAT- Large Hole Image Inpainting

Image completion aims to generate visually appealing and semantically appropriate content for missing areas of a masked image ($IM = I \odot M$). This work proposes a mask-aware transformer (MAT) for large mask inpainting that supports conditional long-range interactions. Additionally, to address the ill-posed nature of the image completion problem, where there may be multiple possible solutions to fill large holes, the approach supports pluralistic generation.

The work proposes:
* A convolutional head to extract tokens
* A transformer body with five stages of transformer blocks at varying resolutions that use multi-head contextual attention (MCA) to model long-range interactions.
* A convolution-based reconstruction module to upsample the spatial resolution of output tokens
* A Conv-U-Net to refine high-frequency details
* A style manipulation module to deliver diverse predictions by modulating the weights of convolutions.

Multi-Head Contextual Attention:
* Handles a large number of tokens and low fidelity in given tokens. * Exploits shifted windows and dynamical masks for non-local interactions. * Computes output as a weighted sum of valid tokens using the attention formula:

$$Att(Q,K,V) = Softmax(\frac{QK^T + M'}{\sqrt{d_k}}) \times V$$

where $M_{ij} = 0$ for valid tokens j, else $-\tau$

Convolutional Head:
* Takes in incomplete image IM and, given mask M, produces 1/8-sized feature maps for tokens.
* It contains 4 convolutional layers, 1 for changing input dimension and 1 for downsampling resolution.
* Used to incorporate local inductive priors in early visual processing, reducing computational complexity and memory cost.

Mask Updating Strategy:
Token validity is initialized by the input mask and automatically updated during propagation. The updating follows a rule that all tokens in a window are updated to be valid after attention as long as there is at least one valid token before.

Transformer Body:
- Processes tokens by building long-range correspondences.
- Contains 5 stages of adjusted transformer blocks with an efficient attention mechanism guided by an additional mask.

Adjusted Transformer Block:
- A proposed variant to optimize masks with large holes.
- Removes layer normalization (LN) and employs fusion learning instead of residual learning.
- Concatenates input and output of attention and uses a fully connected (FC) layer to avoid unstable optimization.
- Observes unstable optimization using general block when handling large-scale masks, attributes to the large ratio of invalid tokens.

Style Manipulation Module
- Weight normalization of convolution layers is changed in reconstruction with additional noise to manipulate the output.
- To enhance the representation ability of noise inputs, the algorithm:
$s_u = \epsilon(n)$
$X' = B \cdot X + (1 - B) \cdot Resize(s_u)$
$s_c = F(X')$

Loss function:
Adversarial Loss = $\min_G \max_D \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[1 - \log D(G(z))]$
Perceptual Loss = $\Sigma_i \eta_i ||\phi_i(x_p) - \phi_i(x)||_1$

4. Experimentation and Datasets

We have in our analysis chosen two domain-related datasets 1. CelebA-HQ and 2. Anime Faces Dataset.

For our analysis, we generate different types of Masks to study the performance of different architectures within different Conditions.

* Noise Masks
Salt Pepper Noise
Gaussian Noise
* Corruption Masks
In these types of masks, a substantial portion of the images are either missing or corrupted.
* Large Masks
A huge chunk of the image is missing.

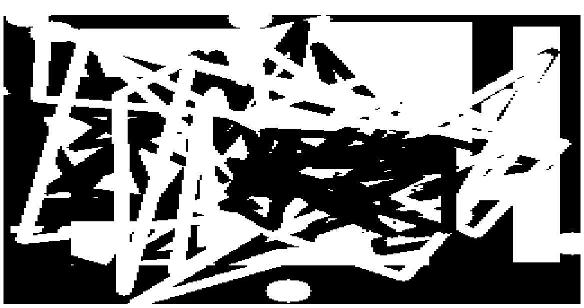

Figure 2. Corrupt

Figure 3. Large

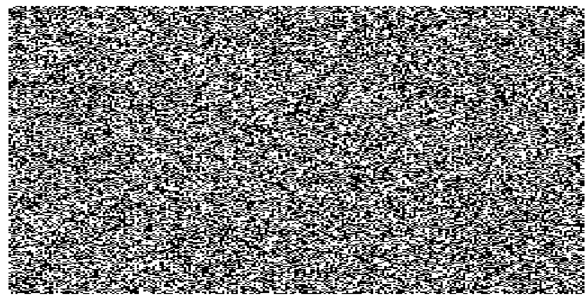

Figure 4. Salt Pepper

CelebA-HQ - Noise Masks

	PDE	Harmonic	MAT	CE-GAN
PSNR	30.60	38.6862	11.1	39.88
SSIM	0.347	0.9826	0.9997	0.9951
MSE	56.55	8.7994	0.0316	0.6121

CelebA-HQ - Corrupt Masks

	PDE	Harmonic	MAT	CE-GAN
PSNR	103.54	45.72	9.1	39.88
SSIM	0.9907	0.9914	0.9937	0.7801
MSE	1.6742	0.5452	0.0232	3.6121

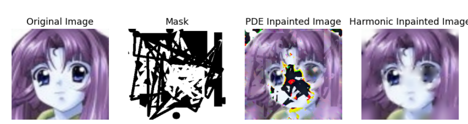

Figure 5. Salt Pepper with MAT

5. Results

We applied both the classical method on the image of a portrait which had a crack, the mask was drawn on the crack and the inpainting methods were used.

Figure 6. Corrupt Mask with MAT

Figure 7. Large Mask with MAT

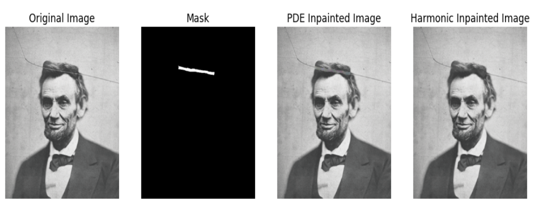

Figure 8. Inpainting using Classical Methods

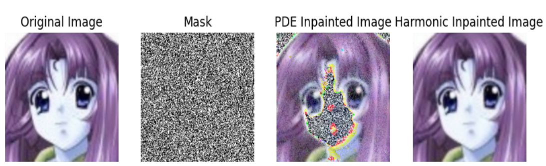

Figure 9. Inpainting using Classical Methods

References

[1] Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. Mat: Mask-aware transformer for large hole image inpainting, 2022. 2

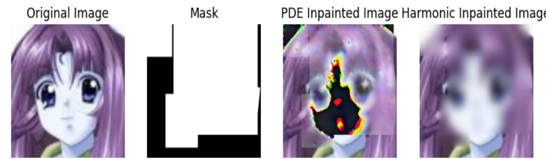

Figure 10. Inpainting using Classical Methods

[2] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting, 2016. 2

[3] Carola-Bibiane Schönlieb. *Partial Differential Equation Methods for Image Inpainting*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2015. 2

[4] Fang Zhang, Ying Chen, Zhitao Xiao, Geng Lei, Jun Wu, Tiejun Feng, Ping Liu, Yufei Tan, and Jinjiang Wang. Partial differential equation inpainting method based on image characteristics. pages 11–19, 08 2015. 2