Project/Thesis No. :

CSE 4000: Thesis/ Project

IMAGE INPAINTING ON IRREGULAR REGION USING GENERATIVE ADVERSARIAL NETWORKS

By

Nishat Tasnim Sithy

Roll: 1807033

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

February , 2024

Image Inpainting on Irregular Region using Generative Adversarial Networks

By

Nishat Tasnim Sithy

Roll: 1807033

A thesis submitted in partial fulfillment of the requirements for the degree of

“Bachelor of Science in Computer Science & Engineering”

Supervisor:

Dr. Sk. Md. Masudul Ahsan
Professor
Department of Computer Science and Engineering
Khulna University of Engineering & Technology (KUET)

____
Signature

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

February, 2024

Acknowledgment

I extend my heartfelt gratitude to the Almighty for His blessings and constant presence. My sincere thanks to my supervisor, Dr. Sk. Md. Masudul Ahsan, Head of the Department of Computer Science and Engineering at KUET, for his unwavering guidance and support. I'm also indebted to my family, friends, and well-wishers for their encouragement and understanding during this journey.

Author

Abstract

Image inpainting refers to filling in missing or corrupted parts of an image with plausible content, creating a completed image that appears seamless and consistent. Image inpainting is a relevant topic today because its application extends across various domains, such as image restoration, editing, and augmentation applications in computer vision. At times, it is difficult to restore an utterly unseen part of an image, and for that purpose, traditional methods require training on many datasets with various masks. Recent advancements in deep learning have shown promising progress in addressing the intricate challenge of inpainting sizable image gaps. While these techniques are proficient in generating realistic structures and textures, they often yield distortions, blurriness, and color discrepancies that don't harmonize with the image context. Given this, opting for a GAN-based approach in image inpainting proves beneficial due to its dual capability: (i) generating novel image content and (ii) leveraging existing features as references during training. This implies that the model learns to draw inspiration from the surrounding context, leading to more coherent and contextually fitting completions. In contrast to conventional methods, GANs introduce the concept of dynamic feature selection, enabling the network to choose pertinent features across various channels and spatial positions adaptively. This thesis presents a novel approach that uses a Conditional Generative Adversarial Networks (cGAN), conditioned explicitly on the masks of images, to target irregularly shaped gaps. This method stands out by employing an innovative technique where the mask of the missing region is dilated, creating an expanded border around the original gap. This dilation guides the inpainting process more effectively by defining a clear boundary for the area to be inpainted. It emphasizes the importance of this border region during the generation process. Thus, the cGAN model is conditioned on both the masked image and the dilated mask, enabling the generator to focus on accurate inpainting of the border regions. This targeted approach ensures that the transitions between the original and generated image parts are smooth and visually coherent, addressing one of the most significant challenges in image inpainting. Currently, the study focuses on natural images to be inpainted. After testing, it can be noted that, although the model is trained to perform on irregular masks, it performs exceptionally

well for regular rectangular masks. The average PSNR of rectangular mask is 28.98, whereas the average PSNR for irregular mask is 19.6.

Contents

List of Tables

List of Figures

CHAPTER I

Introduction

1.1 Introduction

Image inpainting is a technique used to restore parts of images that are missing or damaged. It involves creating new content to fill in the gaps, ensuring that the changes blend in naturally and appear realistic. Image inpainting is becoming increasingly popular due to its wide range of applications and its significant contribution to image restoration, editing, and augmentation in the field of computer vision. The application spans from the restoration of art to the analysis of medical images. Various methodologies have been employed in the field of image inpainting up to the present time. Previously, inpainting heavily depended on the expertise of artists or photo editors to restore the absent regions. The editors utilized basic cloning tools for texture and color synthesis in the digital processes. An early instance of the algorithm was suggested by Bertalmino et al. [1] for employing Diffusion-Based Inpainting techniques to complete the absent areas. Due to the progress in deep learning networks and various generative networks, the process of image inpainting has been simplified. Similar to other deep learning models, a substantial dataset is necessary to train the model and provide it with information regarding texture, color, and pattern. The creation of the filled-in area is frequently dependent on the exposed area to enhance the consistency of the produced image. However, the deep learning model's accuracy is limited in domains with a scarcity of available datasets, such as medical imaging. Hence, post-processing is necessary to enhance the quality of the generated image. This thesis centers on the process of image inpainting using a Conditional Generative Adversarial Network (cGAN), where the model is conditioned on the region of the image that needs to be filled in. The study also presents a comparison between two scenarios: one where the generation prioritizes a border region, and another where the generation does not take the border into account.

1.2 Background Study

The field of image inpainting aims to restore missing or corrupted portions of images while ensuring that the synthesized content seamlessly integrates with the existing scene. The inability of traditional methods to handle complex scenes and generate realistic results makes deep learning approaches necessary to investigate.

While such approaches have shown potential, a state-of-the-art image inpainting model is still required to successfully address the problems of visual quality, semantic coherence, irregular hole handling, and efficiency. The main task is to develop an image inpainting model based on deep learning that can overcome the constraints of current techniques. To support real-time or near-real-time applications, this model should produce inpaintings with remarkable visual fidelity, preserve object structures and contextual relationships through semantic consistency, handle problems caused by large and irregularly shaped holes, and run efficiently. These limitations have inspired the researchers to pursue improved solutions. Consequently, this thesis aims to address these challenges and develop an innovative and comprehensive solution that advances the state of the art in image inpainting and contributes to the broader fields of computer vision, image restoration, and image augmentation.

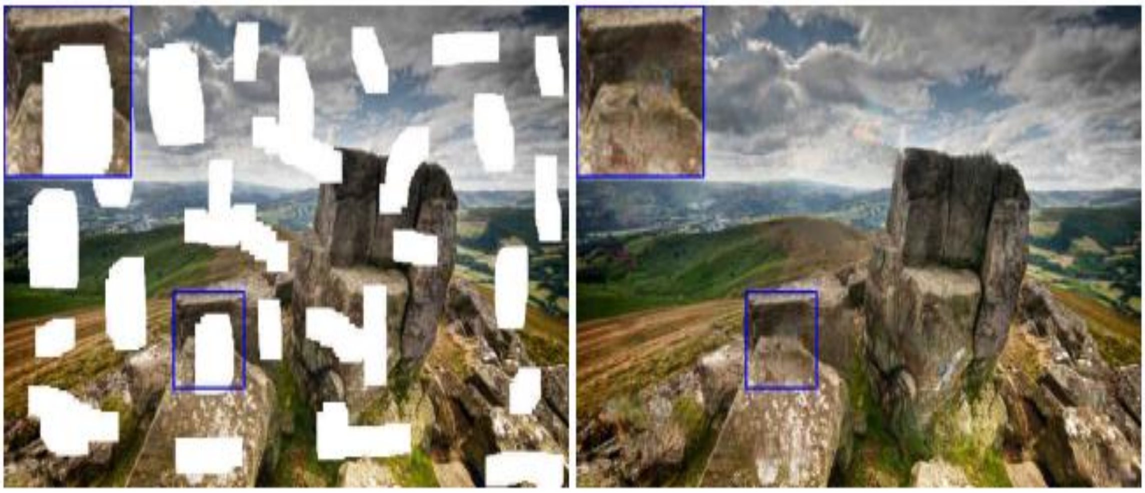

Figure 1.1: Example inpainting results of an existing method on a natural scene is taken from the paper of Lizuka S et al. [2]

The solution will be assessed based on its color accuracy, coherence with the surrounding border, image quality, and ability to restore the details of the input image, in comparison to the most advanced inpainting techniques available. Hopefully, the outcomes will make a valuable contribution to the progress of this discipline.

1.3 Objectives

The central goal of the proposed thesis is to design, implement, and evaluate a state-of-the-art deep learning-based image inpainting model capable of effectively restoring missing or damaged regions in images. The objectives are as follows:

- Employ Conditional Generative Adversarial Networks (cGANs) to restore damaged image regions seamlessly without compromising resolution and study the results.
- Compare the effects of prioritizing a border region of the mask in image inpainting.
- Develop a model to reconstruct larger regions realistically.
- Evaluate the proposed method's performance using a benchmark dataset.

So, the goal is to develop a novel image inpainting that emerges a powerful method and to restore and enhance images seamlessly, where imperfections can be effortlessly mended for a more visually captivating and coherent digital landscape.

1.4 Scope

Image inpainting has a broad scope in various domains. It's used for restoring damaged images, removing objects, video editing, privacy protection, medical imaging, and more. Advancements in AI, like GANs, have improved its realism and applications. This thesis's scope remains within the restoration of natural images, as shown in Figure 1.1. The proposed thesis topic is grounded in the ongoing challenges of image inpainting. While the fundamental idea is well-established, this thesis intends to distinguish itself by combining the model tools and techniques used in the current landscape of image inpainting, including:

1. Deep Learning Frameworks: The model is developed using deep learning frameworks such as TensorFlow or PyTorch, which enable efficient implementation and experimentation.

2. Convolutional Neural Networks (CNNs): CNN serves as the backbone of the deep learning models, enabling it to learn and extract meaningful features and patterns from images to generate high-quality inpaintings.

3. Generative Adversarial Networks (GANs): GANs enhance the realism and quality of inpaintings by simultaneously training a generator and discriminator.

4. Transfer Learning: Pre-trained models like VGG or ResNet is used for feature extraction to make the task more meaningful.

6. Datasets: Model datasets train the model and ensure adaptability to various real-world scenarios.

1.5 Unfamiliarity of the Problem

The unfamiliarity of the problem arises as this topic lies beyond the scope of the standard course curriculum, emphasizing its unexplored nature within academic and practical teaching. Although the basics of image processing have been a part of the academic curriculum, the methodologies or processes related to Image inpainting, such as GANs or CNNs, have not been part of the theoretical or laboratory modules, as a result, selecting this topic as a thesis endeavor presents an innovative and challenging proposition.

1.6 Project Planning

The project planning encompasses several key aspects. A well-defined timeline has been established, aiming to complete the project within a specified duration. Additionally, careful attention has been paid to financial budget considerations to ensure efficient resource allocation. Ethical considerations, as well as potential legal issues about the thesis topic, have been thoroughly examined and integrated into the planning process. This comprehensive approach ensures not only timely completion but also adherence to ethical standards and legal requirements throughout the thesis undertaking.

1.6.1 Project Timeline

The research for the thesis is divided into two phases, each contributing to main goal. The first phase takes place during the first semester, and the second phase is scheduled for completion in the upcoming semester. The target is to have the entire thesis completed by December 2023 to January 2024:

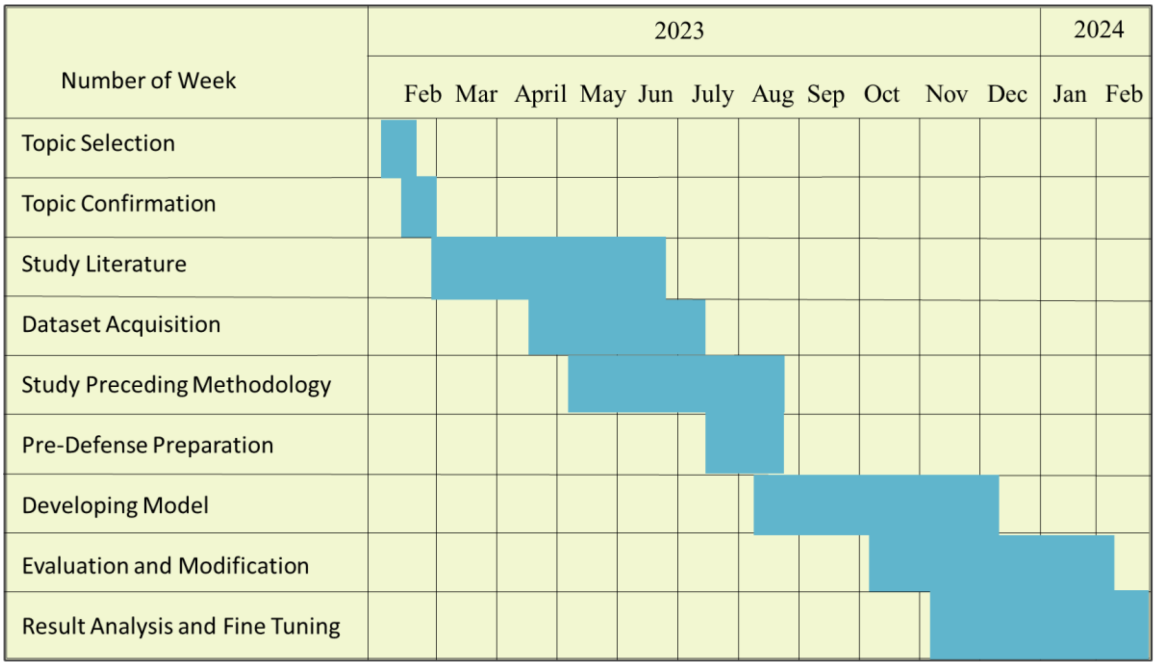

Figure 1.2: Gantt chart of the timeline of the thesis.

1.7 Applications

Image inpainting has diverse applications in fields in real-world applications, including:

1. **Photo Restoration**: Image inpainting helps restore old and damaged photographs by filling in cracks, scratches, and missing parts, preserving cherished memories.
2. **Video Editing**: In video production, image inpainting can remove unwanted objects or artifacts, ensuring a seamless and polished final product.
3. **Medical Imaging**: In medical diagnostics, image inpainting assists in completing missing or corrupted data in medical images, aiding accurate analysis and treatment planning.
4. **Forensic Analysis**: Investigators can use image inpainting to recover obscured details in images, potentially revealing critical evidence in criminal cases.
5. **Heritage Preservation**: In cultural heritage, image inpainting helps conserve artworks and artifacts, reconstructing missing sections to maintain historical integrity.

6. **Virtual and Augmented Reality**: Image inpainting is crucial for blending virtual and real-world elements, enhancing immersion and realism in virtual reality and augmented reality experiences.
7. **Video Game Development**: Game developers use image inpainting to create visually seamless environments, ensuring players' immersion and engagement.
8. **Architectural Visualization**: Image inpainting can help architects visualize completed structures by filling in details, aiding in design and presentation.
9. **Image Compression**: In data storage and transmission, image inpainting techniques contribute to efficient compression methods, reducing file sizes while maintaining image quality.
10. **Content Generation**: Image inpainting is utilized in generating realistic textures, landscapes, or objects in computer graphics and design applications.
11. **Environmental Monitoring**: In remote sensing, inpainting helps reconstruct missing data in satellite images, improving our understanding of environmental changes.
12. **Criminal Investigation**: Law enforcement agencies can utilize image inpainting to enhance and restore visual evidence, assisting in criminal investigations.

These are some common applications of image inpainting, required in practical life.

1.8 Organization of Report

The thesis report is organized into five chapters. Each chapter provides us with a clear perception of how each stage is planned and executed, giving an overview of the overall thesis.

Chapter I presents the background of image inpainting, along with the problem statements, and what objectives are planned to achieve with this thesis. The planning of the thesis is also included.

Chapter II provides an insight into current advancements in the field of image inpainting. Along with the comparative discussion between their methods and architecture.

Chapter III provides an overview of the methodology that has been followed throughout the thesis.

Chapter IV provides an idea of the technical details, including the experimental setup required for conducting the thesis, along with implementations, results, and a comparative evaluation. It also discusses the financial budget planning for the thesis and its achieved objectives.

Chapter V discusses a summary overview of the work, alongside the ethical, safety and legal considerations of thesis. This includes the socio-economic impact of the thesis as well.

Chapter VI addresses the complex engineering problems and applications of the thesis work.

Chapter VII summarizes the thesis, and discusses the limitation and future scopes

1.9 Conclusion

In the upcoming sections a detailed exploration of every step this thesis research has been explained. Through in-depth discussions and thorough analysis, the intention is to provide a clear and comprehensive view of the processes, challenges, and outcomes that have played a pivotal role in the journey of exploring image inpainting techniques using cGAN.

CHAPTER II

Literature Review

2.1 Introduction

Image inpainting, a pivotal domain within the realm of computer vision and image processing, has garnered substantial attention for its ability to seamlessly restore missing or damaged portions of images. This technique holds paramount importance across diverse applications, including photo restoration, video editing, medical imaging, and forensic analysis. Within the literature, a plethora of methodologies have been explored, each offering distinct approaches and techniques to tackle the complex challenges posed by inpainting irregular holes, preserving semantic coherence, and achieving high-quality, visually appealing results. In this context, a comprehensive literature review aims to navigate through the diverse landscape of image inpainting techniques, highlighting their strengths, limitations, and contributions to the broader field of image processing. The following sections delve into the key methodologies and architectures employed by various models, drawing insights from comparative analyses to illuminate the advancements and potential directions within this dynamic area of research.

2.2 Relevant Terminology

This section briefly describes some necessary theoretical concepts related to the image inpainting process for a better understanding of the reader. In this thesis, GAN-based methods are the main consideration. To use a GAN-based model, some additional concepts need to be understood. A brief introduction to those concepts will be given in this chapter.

2.2.1 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN), first introduced in paper by Lecun et al. [3] is a deep learning architecture designed for processing structured grid data, such as images or sequences. It employs layers of learnable filters that automatically learn and extract features from the input data. These filters perform convolutions across the input, enabling the network to capture spatial hierarchies of features.

In image inpainting, CNNs can be employed to learn the relationships between known image regions and their corresponding context. By training on examples of images with missing parts and their complete versions, CNN can learn to predict the missing content based on the surrounding context.

2.2.2 Partial Convolution

Partial Convolution is a technique used in image processing and computer vision tasks where only valid parts of the convolutional kernel are used for areas that have known information (non-zero values). In areas with missing or masked information (zero values), the convolutional kernel is adjusted to account for the reduced receptive field. Partial Convolution is particularly useful in image inpainting , as shown in [4] because it allows the convolutional operation to adapt to the available information. When inpainting an area, Partial Convolution ensures that only valid parts of the convolutional kernel are used for known image regions, while the kernel adjusts to account for the missing or masked parts. This prevents the inpainting process from introducing artifacts in areas with incomplete information. Interested readers can find more in the paper by [5]

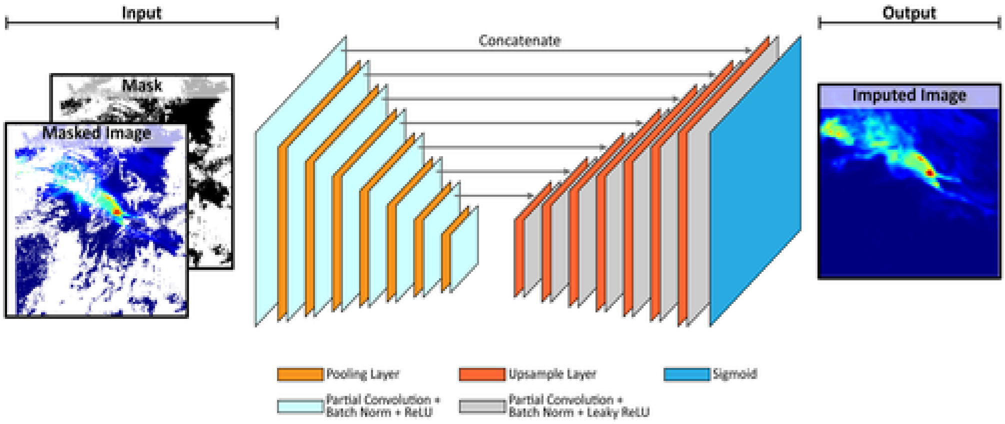

Figure 2.1: Schematic of the partial convolutional neural network model by [6]

2.2.3 Gated Convolution

Gated Convolution is a convolutional operation that combines the outputs of two separate convolutional filters using gating mechanisms. This allows the network to selectively pass information from different filters, enhancing the network's capacity to capture complex patterns and relationships in the data. Gated Convolution can enhance the inpainting process by enabling the model to selectively pass relevant information from different convolutional filters. This can help the model learn to fill in missing regions more accurately, considering both local and global contexts. By controlling the flow of information through gating mechanisms, the model can prioritize the most suitable features for inpainting. Interested readers can find more in the paper by Yu et al. 2018 [7]

2.2.4 U-Net Architecture

The U-Net architecture is a deep learning framework for semantic segmentation tasks in computer vision. It consists of a contracting path that captures context and a symmetric expanding path that produces segmentation maps. The architecture's unique "U" shape aids in capturing both low-level and high-level features, making it effective for tasks like biomedical image segmentation. The U-Net architecture is well-suited for image inpainting due to its contracting and expanding paths. The contracting path captures the context and structure of the image, while the expanding path generates the inpainted content. This allows the model to utilize both local and global features for accurate inpainting. The symmetrical structure of U-Net ensures that the generated content maintains spatial consistency with the surrounding region. Interested readers can find more about at paper by [8].

Figure 2.2 gives an architectural view of U-Net architecture.

2.2.5 Hourglass Architecture

The Hourglass Architecture is a neural network design commonly used for tasks like human pose estimation and keypoint detection. It involves a sequence of down-sampling and up-sampling layers, forming a symmetrical "hourglass" shape. This architecture allows for the capture of both local and global features, making it robust for tasks that require precise spatial localization. Hourglass Architecture's ability to capture both local and global features makes it effective for image inpainting. By downsampling and then upsampling, the model can capture details at different scales, helping to fill in missing areas with coherent and contextually relevant content. Interested readers can find more in the paper [5]

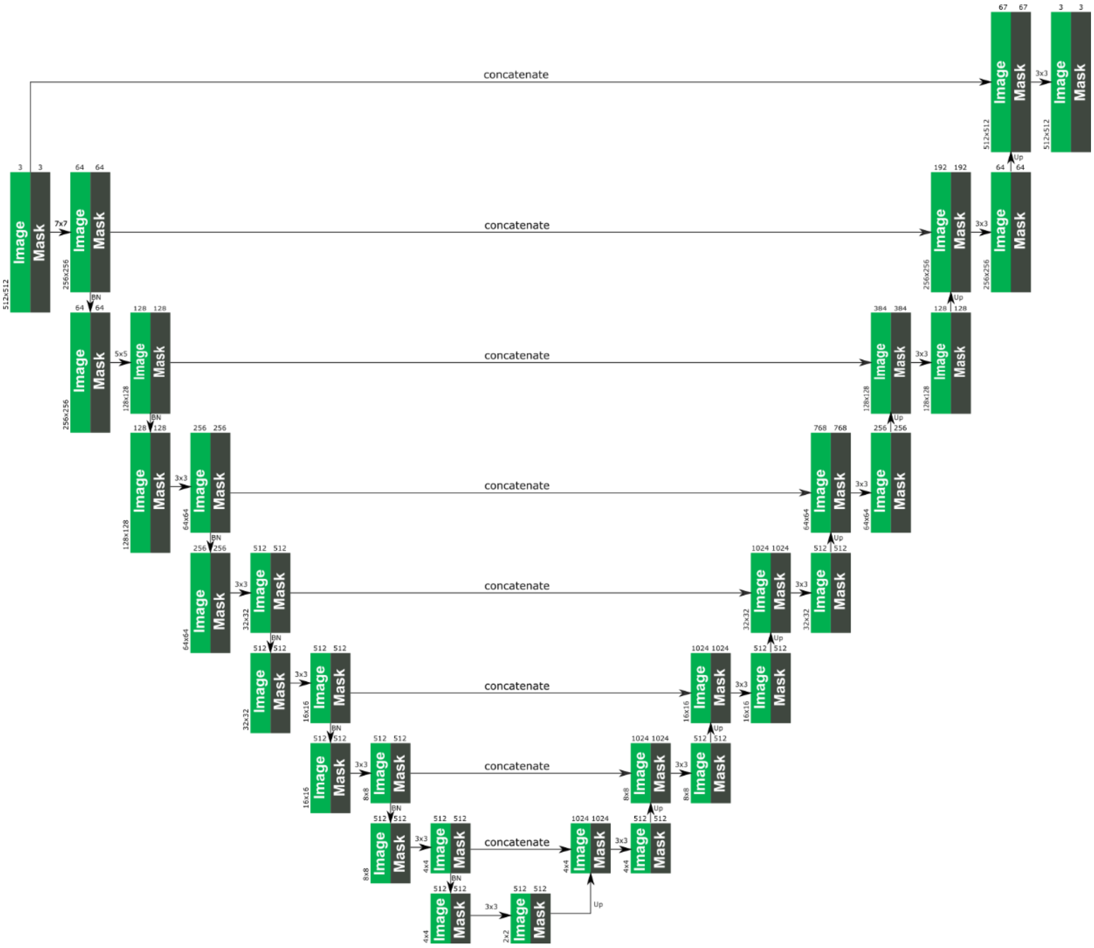

Figure 2.2: Architecture of U-Net from [8]

2.2.6 Encoder-Decoder Architecture

An encoder-decoder architecture is a type of neural network design that consists of two main components: an encoder and a decoder. The encoder takes the input data and compresses it into a lower-dimensional representation (latent space). The decoder then takes this encoded representation and reconstructs the original data. This architecture is commonly used in tasks like image denoising, image inpainting, and image-to-image translation. In image inpainting, the encoder-decoder architecture is used to generate plausible content for missing or damaged regions within an image. The architecture leverages the encoder to capture relevant features from the known parts of the image, and the decoder then generates the missing content based on these features.

2.2.7 Generative Adversarial Network (GAN)

A Generative Adversarial Network (GAN) is a framework that consists of two neural networks, a generator and a discriminator, trained simultaneously through a competitive process. The generator tries to create data that is indistinguishable from real data, while the discriminator aims to classify accurate and generated data correctly. This adversarial training process results in the generation of high-quality synthetic data. GANs can be used for image inpainting by training a generator to create plausible content for the missing regions. The discriminator evaluates the realism of the inpainted content, providing feedback to the generator. This adversarial process encourages the generator to produce realistic and visually consistent inpainted images. Interested readers can find more at [7], [8].

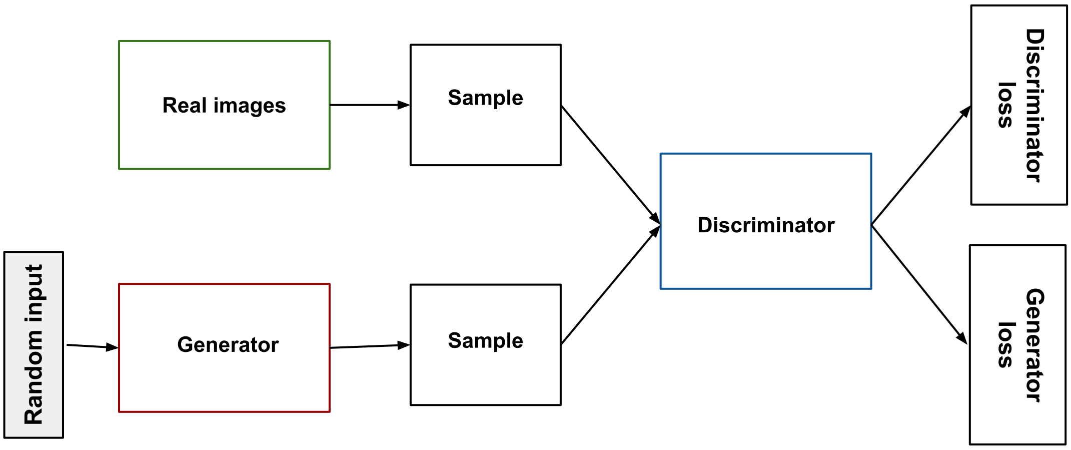

Figure 2.3: An architectural overview of GAN

2.2.8 Conditional Generative Adversarial Networks

The conditional Generative Adversarial Network is a special type of GAN model, which allows the generation of synthetic data samples to be conditioned on additional input information. In a cGAN, the generator network takes an additional input, such as an image label or an image, as input and produces synthetic data samples that are conditioned on that input. When it comes to image inpainting, the condition that works here is a binary mask, that guides the generator to produce a specific part of the image corresponding with the surrounding part of the image. The discriminator then tries to discriminate between the real and fake data for the applied condition. Once the training is over, the generator is able to produce realistic images based on the similar types of conditions. Some other application of cGANs include image colorization, style transfer, text-to-image synthesis and so on. One of the prominent papers on cGAN can be found on the literature by Mirza et al. [9]

2.2.9 Activation Functions

An activation function is a fundamental component of a neural network, specially a generative adversarial network. It introduces non-linearity into the computation of the network and determines the output of a neural node, given a set of inputs. The main task of an activation function is to transform the input signal into an output signal for the next layer. Some important activation functions are:

- **ReLU (Rectified Linear Unit):**
  It outputs the input directly if it is positive, otherwise, it outputs zero. It helps to reduce the vanishing gradient problem, allowing models to learn faster and perform better. Although it may lead to dying ReLU problem, meaning the neurons output zero for all inputs and gradients do not flow through neurons anymore. It is defined by:

  $$f(x) = \max(0,x)$$

- **LeakyReLU:**
  It is a variant of ReLU designed to address the dying ReLU problem. It prevents the neurons from dying by allowing a small, non-zero gradient (α) when the input is negative.
- **Tanh (Hyperbolic Tangent):**
  This outputs values in the range (-1,1). It is often used in hidden layers of neural networks as it centers the output, making it zero-mean, which can help with the convergence during the training. Since it outputs in the range of (-1, 1) it can be more effective then LeakyReLU.

Each of this functions have their own user cases and they are chosen based on the requirement of the task and the goal of the neural network. Interested reader can find more about activation layers in [10], [11], [12]

2.3 Related Works

The comprehensive review of the existing literature in the field of image inpainting reveals a distinct bifurcation that organizes the various approaches into two overarching categories: "Machine-Learned Inpainting Approaches" and "Structure-Based Inpainting Approaches." This categorization serves as a valuable framework to navigate through the diverse methods employed for image restoration and completion. The first category encompasses conventional methods that rely on diffusion or patch-based methodologies, utilizing basic features. The second category is characterized by a more modern approach involving learning-based techniques. An example of this entails training deep convolutional neural networks to predict and fill pixels within areas requiring completion within an image.

2.3.1 Structure-Based Inpainting Approaches.

The earliest techniques included manual inpainting methods that were solely dependent on the skill of the artists or photo editor to reconstruct any missing areas. It involved simple algorithms for texture synthesis and cloning tools. One algorithm that needs to be mentioned is Diffusion-Based Inpainting by Bertalmio et al. [1] which propagates linear structures in the hole region, mimicking the lines of equal color intensity. This method worked effectively for small scratches and gaps.

Next it is important to mention the use of Patch-Based or Fragment-Based image completion by Drori et al. [13]. This method copies and pastes small patches from image itself to fill in larger gaps. Helpful for texture synthesis, where the texture is repetitive.

As described in the paper by Yu et al. [8] and other sources, traditional diffusion or patch-based approaches typically employ variational algorithms or patch similarity to propagate information from background regions into the holes. These methods work effectively for stationary textures but have limitations when dealing with non-stationary data, such as natural images. Ulyanov et al. [14] proposed a method in which a randomly initialized neural network serves as a handcrafted prior for standard inverse problems like image inpainting. Deep convolutional networks are employed as handcrafted priors for image generation and restoration. This approach eliminates the need for training on a dataset, bridging the gap between learning-based and learning-free methods that rely on handcrafted priors. However, this method's two primary limitations are its slow execution and its inability to match or surpass the results of problem-specific techniques in practical applications. Due to its non-trainable nature, this method is not expected to perform well on "highly semantic" large hole inpaintings, such as face inpainting.

2.3.2 Machine-Learned Inpainting Approaches.

Jiahui Yu et al. [8] propose an alternative generative image inpainting approach utilizing contextual attention. The authors present a unified feed-forward generative network equipped with a novel contextual attention layer for image inpainting. This network consists of two stages: (i) a simple dilated convolutional network trained with reconstruction loss to approximate missing contents, and (ii) a second network that takes the coarse prediction as input and produces refined results. The core concept of contextual attention involves using features from known patches as convolutional filters to process generated patches. The method employs mirror padding for convolution layers, ELUs as activation functions, and output filter value clipping. The entire network is trained end-to-end using reconstruction losses and two Wasserstein GAN losses. Notably, this approach demands a substantial amount of training data to effectively learn to inpaint various image types, which could pose challenges when applying the method to domains with limited training data, such as medical imaging. Furthermore, the method's performance suffers from irregularly shaped holes, rendering it unsuitable for practical applications.

To overcome the limitations of the aforementioned methods, Liu et al.[6] proposed employing partial convolution layers, comprising a masked and re-normalized convolution operation followed by a mask update step. The model adopts a UNet-like architecture by replacing all convolutional layers with partial convolutional layers and utilizing nearest neighbor up-sampling during the decoding stage. The authors suggest that this approach robustly handles holes of varying shapes, sizes, locations, and distances from image borders. However, it may struggle with sparsely structured images and larger holes and can lead to errors when holes are situated near the image's border.

For advanced creative editing and object removal, Jiahui Yu et al. [7] introduced a generative image inpainting system founded on gated convolution, learned from an extensive image dataset. The authors also introduced a patch-based GAN loss known as

SNPatchGAN [reference to be added]. Unlike the PartialConv [7] method that employs a U-Net architecture, this approach employs a simpler encoder-decoder network. The inpainting network is trained end-to-end and accommodates free-form holes at arbitrary locations. Nevertheless, it falters when the hole is near the image boundary—a critical scenario in real-life photograph restoration.

As one of the most useful methods, Yeh et al.[15]propose a semantic image inpainting method that generates missing content by conditioning on available data. The process seeks an encoding of the corrupted image closest to the image in latent space, reconstructing it using a generator. This method does not necessitate masks for training and can be applied to arbitrarily structured missing regions during inference. The model's experiments utilize a DCGAN [Reference to be added] architecture. The authors propose that this approach yields images with sharper, more realistic edges and, due to its ability to learn training data representations, can predict meaningful content for corrupted images. However, the current GAN model within the proposed method performs well for simple structures like faces but is too limited to

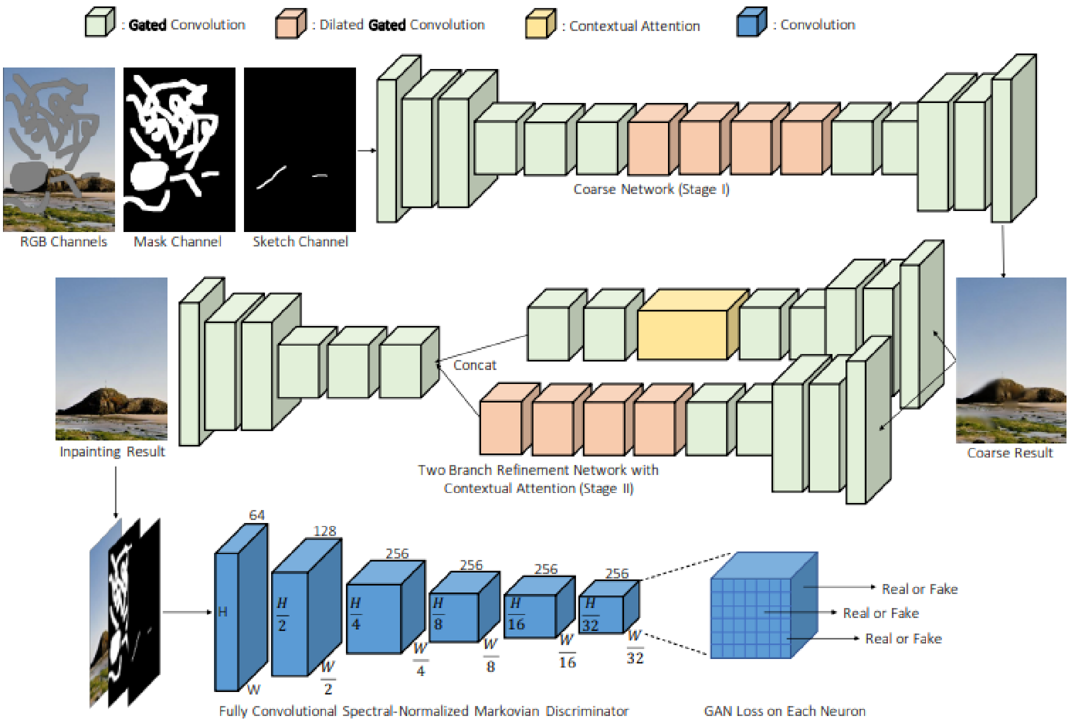

Figure 2.4: Overview of the framework with gated convolution and SN-PatchGAN

effectively represent complex real-world scenes. Enhanced generative models could readily improve the method's performance.

Lastly another GAN based approach was observed in damaged font repair, by Liu et al. n. d. [16]that uses conditional generative adversarial networks and inputs the broken calligraphy works into the generator as x and a random vector z in SCCGAN. The random variable z learns the data distribution in the reference image y, generating b z (the damaged region in x) and the damaged calligraphy composition to achieve the repair effect. The paper also uses the VGG network model pre-trained on ImageNet to learn the writing style of Chinese characters.

2.4 Discussion

Among the reviewed models, each offers distinctive approaches and strengths in image inpainting. The Deep Image Prior model by Ulyanov et al. [14] leverages untrained networks as priors, excelling in text removal and structured inpainting. In contrast, the Generative Image Inpainting with Contextual Attention model by Yu et al. [8]employs parallel encoders and contextual attention to reduce artifacts and prioritize context-driven Here are the results for the Image Inpainting for Irregular Holes Using the Partial Convolution model by Liu et.al. [6] stands out for its effectiveness in handling varied hole shapes and sizes through its UNet-like structure.

The Free-Form Image Inpainting with a Gated Convolution model by Yu et al. [7] introduces a user-guided creative editing paradigm with gated convolutions. Meanwhile, the Semantic Image Inpainting with a Deep Generative Model by Yeh et al. [15] excels in achieving pixel-level realism for large missing regions. However, the latter's performance hinges on the generative model and training. In comparison, the other models demonstrate robustness across tasks, albeit with varying training intricacies.

Choosing the appropriate model depends on the specific task's requirements. For structured inpainting, the [8], [14]might be preferable. For irregular holes, the [6]offers an advantage, while creative editing benefits from the model [7]. Semantic inpainting, particularly with large missing regions, aligns well with [15] Evaluating the trade-offs between model performance, training complexities, and intended tasks is crucial in making an informed decision.

In the method proposed by Liu et al. focuses mostly on the style of the font as well as the content of the images. Thus, the double discriminator does a great job in generating a content with similar style. This characteristic is very important for image inpainting since not only the generated image has to be of good quality it also need to be coherent with the style of the original image to make the output appear realistic.

2.4.1 Research Gap Solution

The literature review of this topic has been a great guide for identifying the research gap. This helped to come up with the new idea and solution. The idea is to come up with a method to identify and color and texture of the surrounding region, and generate the masked area. To make sure the generated images appear seamless, a border map is used that identifies the surrounding region around the mask and then a conditional generative adversarial network is used to condition the generation on the mask.For a better understanding, the comparison between the method and architecture is shown in the following table 2.1. Then the general model, architecture and limitations of this thesis has been also included in the table for a comparative study. It can be seen the thesis is not directly based off from any of the aforementioned research, rather inspired from the limitations and research gap from the papers.

Table 2.1: Comparative study among the related works.

Models proposed by	Method	Architecture	Remark	Scope for Improvement
Ulyanov et al	the randomly initialized generator network	Hourglass (decoder-encoder) with optional skip connections.	- Limited for non-stationary data - May not handle complex structures well	- automation of hyperparameter tuning
Yu et al	contextual attention	Generative network with contextual attention.	- Requires substantial training data - Might struggle with irregular hole shapes	- refining the attention mechanism - handle irregular hole

Models proposed by	Method	Architecture	Remark	Scope for Improvement
Liu et al	partial convolutions	UNet-like architecture with partial convolutions	- May fail for sparsely structured images and large holes - Errors near image borders	- improve for sparsely structured images and larger holes.
Yu et al	gated convolutions and SN-PatchGAN loss.	Coarse and refinement networks with gated convolutions.	- Limited performance near image boundaries - May not effectively handle complex scenes	- optimization of balance between image quality and computational efficiency
Yeh et al	trained generative model	Generative model with encoder, decoder, discriminator	- Limited by the size of the GAN model in representing complex scenes	-improvement of representation of complex scenes and structures
Liu et al	GAN based method with double generator	UNet based generator and the discriminators are used to calculate style loss and content loss	-Limited for repairing damaged font style. - Domain-specific, can not be targeted for a broader domain with multiple styles,	Font strokes tend to merge for dense fonts, and large areas of content loss might result in blank spaces without any strokes.
The thesis	A conditional GAN	The generator uses U-Net based architecture and the cGAN is conditioned on the mask and border of mask	-Only targets the natural scenes, and is trained on one such dataset. - Fails to generate realistic structure for larger area	Improvement can be made to generate realistic texture and structure in larger region.

2.5 Conclusion

In conclusion, the comprehensive exploration of the literature reveals a diverse landscape of image inpainting techniques. From non-parametric methods like exemplar-based approaches to parametric deep learning models, each model offers distinct advantages and addresses specific challenges in the realm of image restoration. While some models excel in structured inpainting, others demonstrate prowess in handling irregular holes or achieving pixel-level photorealism. The reviewed literature underscores the ongoing evolution of image inpainting, with models showcasing advancements in contextual attention, gated convolutions, and deep generative frameworks. As this thesis delves deeper into this landscape, it aims to synthesize and extend these methodologies, further refining state-of-the-art in image inpainting and contributing to the broader field of computer vision.

CHAPTER III

Methodology

3.1 Introduction

The proposed methodology sets the foundation for advancing the field of image inpainting by amalgamating insights from diverse models while introducing novel elements. Building upon the strengths of existing approaches, the methodology is designed to create a more versatile and efficient image inpainting model that bridges the gap between structured and free-form inpainting challenges. As previously mentioned, in this thesis work, the image generation of the masked region is done using Conditional Generative Network. This architecture consists of two blocks: a generator and a discriminator. Once the training is over, the generation can be done using the generator block. This chapter includes the detailed methodology, problem design and analysis, framework and flowchart, and different architectures.

3.2 Detailed Methodology

Before the implementation and designing the architecture, the thesis needs to be divided into small problems, and the solution must be designed accordingly. The main objective is to tailor a cGAN framework for the task of image image inpainting,

3.2.1 Problem Design and Analysis

The process of generating an occluded region from a masked image by using a cGAN can be divided into the following stages:

1. Dataset collection
2. Binary Mask Generation
3. Pre-processing of Dataset
4. Calculation of the loss function

5. Training the generator
6. Extracting the trained generator

Now the generator can be used for inpainting the masked region of an image. A visual representation of the workflow is given below:

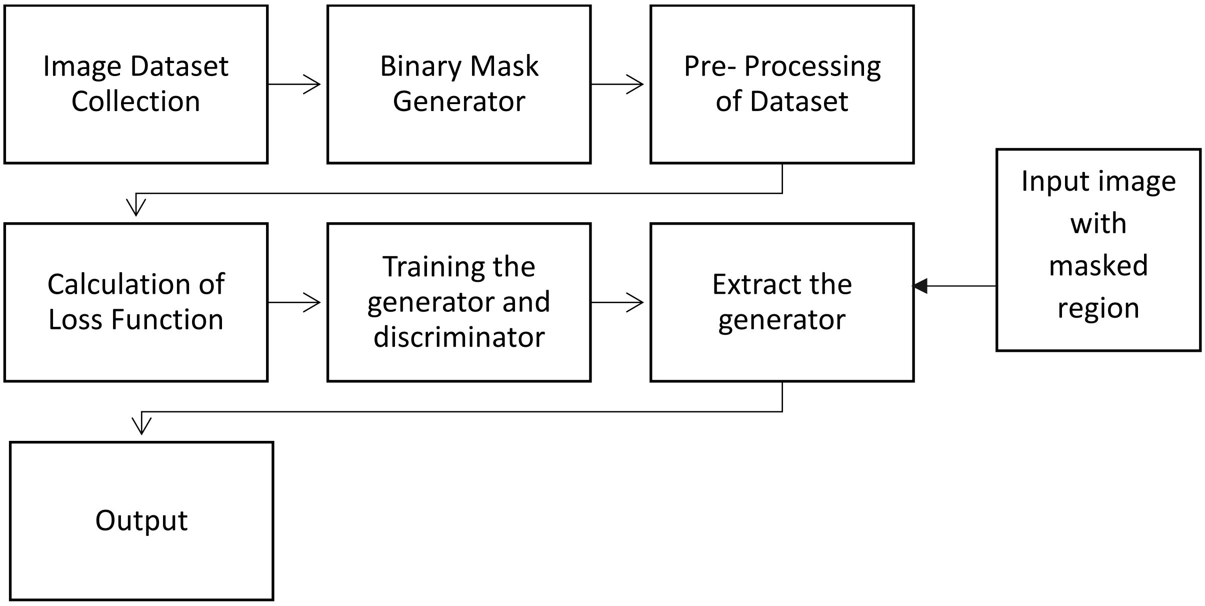
Image Dataset Collection
Binary Mask Generator
Pre- Processing of Dataset
Input image with masked region
Calculation of Loss Function
Training the generator and discriminator
Extract the generator
Output

Figure 3.1: Workflow of image inpainting using CGAN

The before the dataset is fed to the generator for training, the following steps are followed for pre-processing:

- **Image-Mask Pairing**: Each image is paired with a corresponding mask. These masks specify areas within the images that need inpainting.
- **Transformations**: If provided, transformations are applied to both images and masks to potentially resize, crop, normalize, or augment the data in preparation for training.
- **Mask Processing**: Masks are processed to ensure they are binary and then dilated to increase the size of the masked areas slightly. This helps distinguish between the areas to inpaint and the surrounding pixels.
- **Weight Map Creation**: A weight map is generated, giving higher importance (or weight) to the pixels around the border of the masked areas. This focuses the model's attention more on these border regions during the inpainting process.

- **Image Masking**: The original images are masked, setting the pixels of the masked regions to zero, preparing them for the inpainting task.
- **Weighted Image Preparation**: The masked images are multiplied by the weight map, emphasizing the importance of border pixels in the learning process.

3.2.2 Loss Calculation

To calculate the loss for the training, 2 losses were considered. That includes Adversarial Loss, Per Pixel Loss.

- **Reconstruction Loss:** This loss measures the pixel-wise difference between the ground truth I_{gt} and output image I_{out} . This promotes the fidelity to the original image. The equation of the loss function:

  $$L_{rec} = || I_{gt} – I_{out} ||$$

- **Adversarial Loss:** Adversarial loss is calculated using the discriminator. The discriminator generates a real or fake verdict on the generated and ground truth images. Their logarithmic difference is known to be the adversarial loss

  $$L_{adv}= \log D(I_{gt})+\log(1-D(I_{out}))$$

3.2.3 Detailed Architecture of the cGAN:

The cGAN architecture can be broadly divided into the Generator and the Discriminator architecture. To train the entire model, both the generator and the discriminator need to be trained separately in the training loop. The model is conditioned on the mask, and so it can be said that the mask guides the entire training process. So, the generator takes two inputs: The masked image and the corresponding mask.

Before the masked image is passed to the generator, a weight map needs to be created to emphasize the border around the masked region. This is done by performing dilation to the original mask. Dilation enlarges the masked region. Now when the original mask is subtracted from the dilated, a border region is found. This border region is prioritized during the inpainting process. During convolution, the border region has more weight than the other pixels. This process makes sure the generated region is coherent with the surrounding pixel.

At the beginning of the training, the discriminator is trained before the generator. In the discriminator training phase, a masked image is first fed to the current generated model, and the generator returns an inpainted image labeled as a “fake image.” And the corresponding complete image is labeled a “real image.” The discriminator tries to distinguish between the two, measured using “Discriminator Loss.” The discriminator loss is based on how well the discriminator performs in discriminating between the two.

Then comes the training of the generator. While training the generator, the aforementioned losses are calculated, and the total loss is a weighted sum of the individual loss components. Each type of loss contributes to different aspects of the inpainting quality, and their weights determine their relative importance. During the training, total loss is minimized using backpropagation. The gradients of the loss with respect to the model parameters are computed, and those gradients are used to update the model’s weights. This process is repeated over many iterations, with the model gradually improving its inpainting performance. Figure 3.3 shows the overall framework of the cGAN model for better understanding.

3.2.4 Discriminator Architecture

The discriminator is designed as a series of convolutional blocks that incrementally downsamples the input images, distinguishing the real images from the inpainted images, generated by the generator.

.

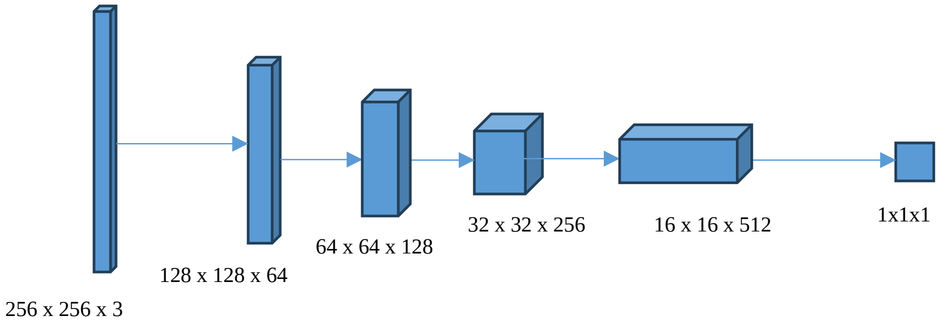

Figure 3.2: The discriminator architecture of cGAN for image inpainting.

The initial block contains a convolutional layer using LeakyReLU activation but not batch normalization. The rest of the block increases the number of filters while continuing to downsample the image, with a constant convolutional layer structure followed by batch normalization and LeakyReLU activation. This enables the network to extract complex and abstract features from input images. The discriminator reduces the feature map to a single scaler output in the final convolutional layer. This output represents the discriminator's assessment of the image's authenticity. The following diagram shows the architecture of the discriminator.

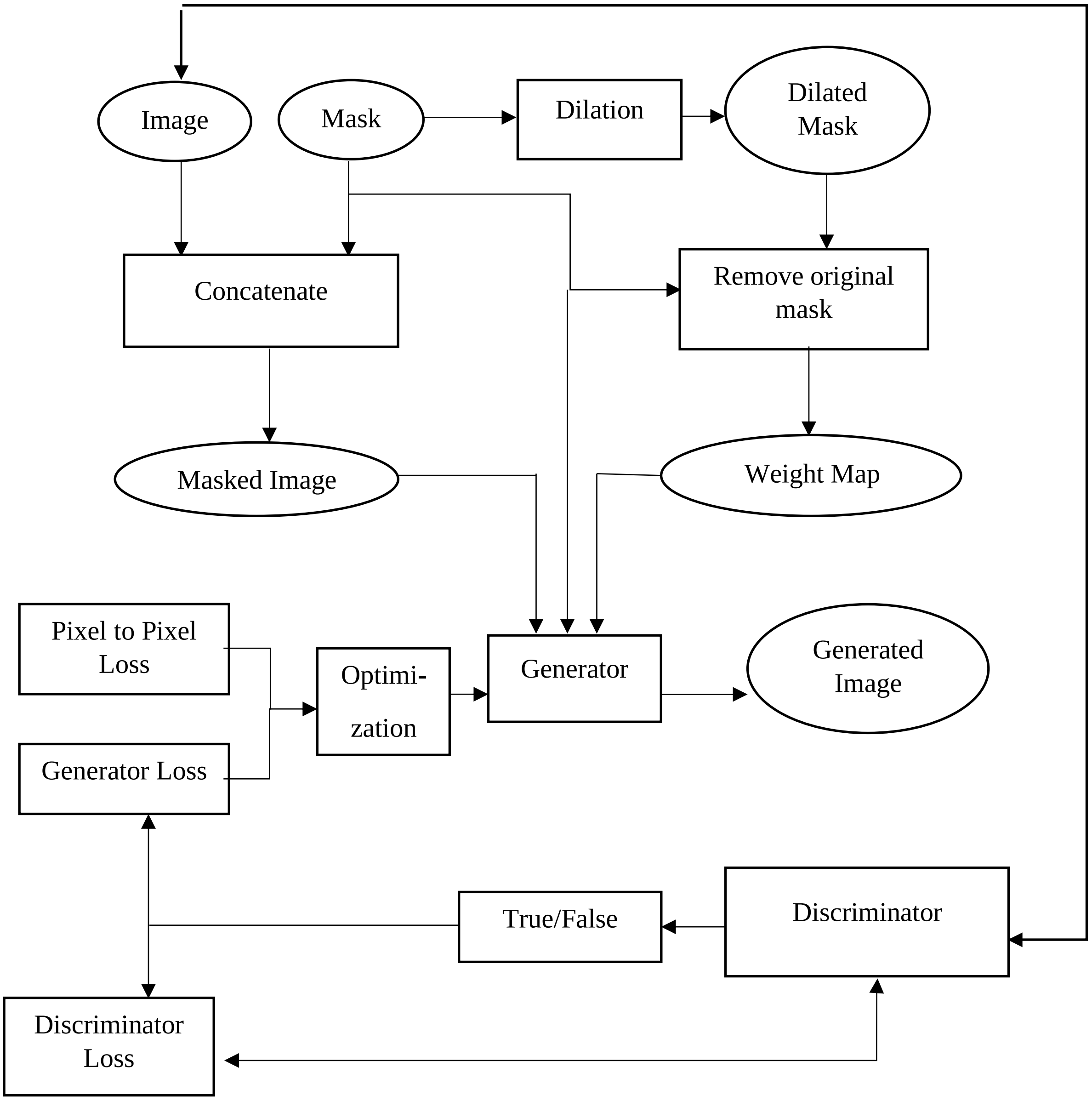

Figure 3.3: The general architecture of the cGAN network for image inpainting.

3.2.3 Generator Architecture

The generator employs a UNet based architecture. It is characterized by its encoder decoder structure with skip connections. The encoder part which consists of 5 layers, is for downsampling. The downsampling layer is for capturing basic features from the input image while reducing its spatial dimensions. Each subsequent layer doubles the number of filters while further downsampling the spatial dimensions. The decoder part also consists of 5 layers. The first layers progressively upsamples the feature maps while halving the filters at each layer. The skip connections from the corresponding encoder layers are concatenated to the upsampled feature maps, which introduces the lost spatial information.

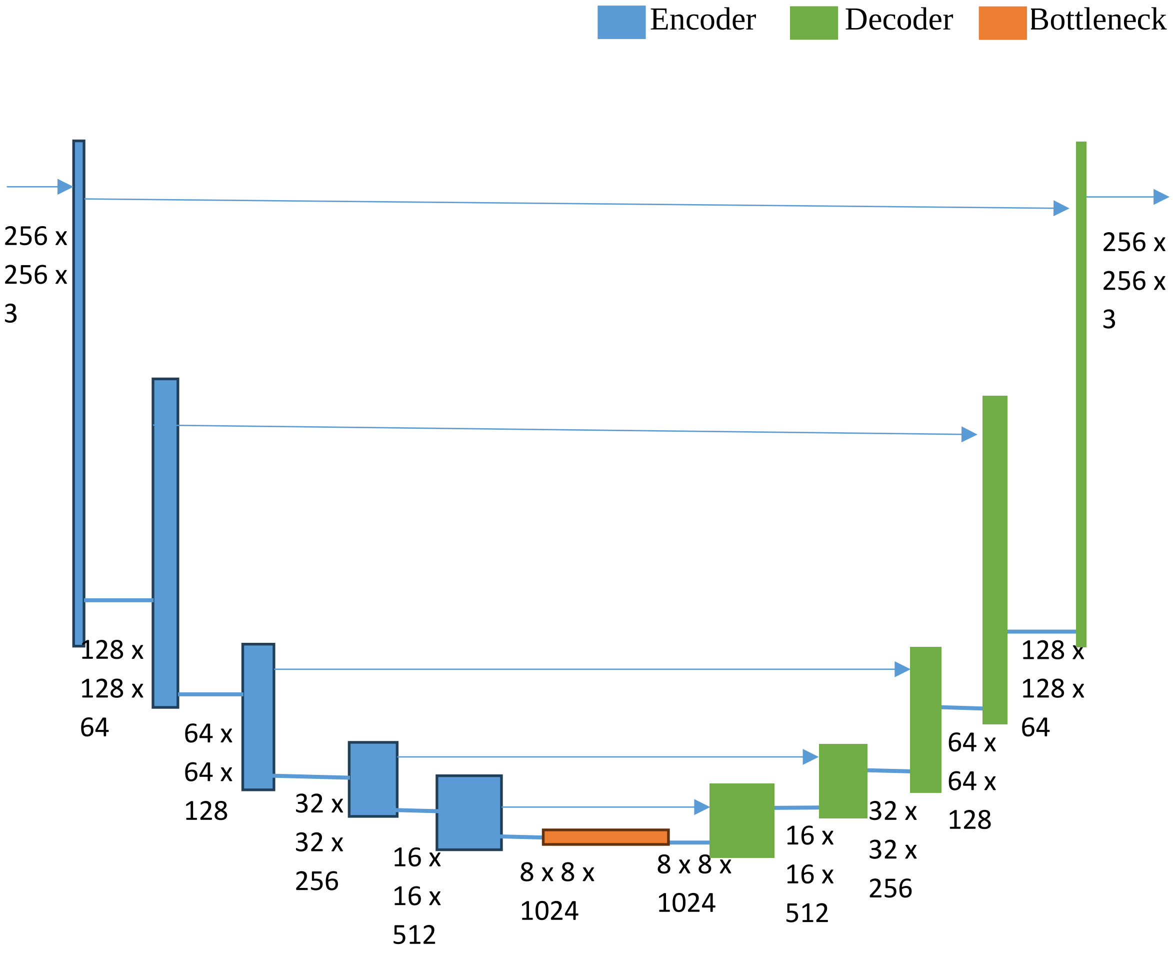

Fig 3.4: Generator architecture for the image inpainting model

3.4 Conclusion

In conclusion, the methodology represents a synthesis of the best elements from established image inpainting models, along with innovative modifications tailored to overcome their limitations. By integrating the strengths of different approaches, the methodology tries to create an adaptable and high-performing image inpainting model that excels in producing contextually coherent and visually appealing inpainting. The U-Net based approach consists of an encoder path followed by a decoder path, which at first reduces the spatial resolution while increasing the number of channels and then upsamples the features back to the original resolution. This structure allows the network to capture the local and global context information. The skip connections facilitate the flow of information from early layers to later layers, which enables the layers to preserve better details during inpainting.

CHAPTER IV

Implementation and Results

4.1 Introduction

In the implementation phase, the process begins with curating a diverse dataset of images featuring varying missing regions. Progress is made through data preprocessing, encompassing mask creation and augmentation. In this section, the experimental setup, the implementation of the methodology, the qualitive and quantitative result and the analysis of the result is being discussed. Finally, the achieved objectives are compared with the initial objectives of this thesis.

4.2 Experimental Setup

The optimal setup proposed by the authors:

Hardware Setup:

1. GPU: NVDIA GeForce GTX 1080
2. Processor: Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz 2.71 GHz
3. Memory: 8Gb VRAM
4. RAM: 32GB RAM 2133MHz

Online Setup:

1. GPU: NVIDIA Tesla P-100-PCIE-16GB, NVIDIA T4(x2)

Software Setup:

1. Operating System: The model has been trained on Windows10
2. Programming Language: python v3.9.18
3. IDE: Spyder v5.0.0
4. Environment: Anaconda or Kaggle

5. Deep Learning Frameworks and Libraries: Many deep learning frameworks have been used to develop the model and the thesis, following are some important frameworks that have been used:
    - Tensorflow v2.9.1
    - Pytorch-cuda v11.7
    - Opencv v4.6
    - Numpy v1.5.0
    - Torch v1.12.1+cu113
    - Torchvision v0.13.1+cu113

4.3 Implementation

It is known that, every deep learning model requires to be trained on a large dataset. So, the first step of implementation is to collect a large dataset. As there is no dedicated dataset necessary to train the model, a large dataset of diverse image is needed. Along with that, a large amount of mask dataset is also required to synthetically distort the original image. The implementation process is discussed below:

4.3.1 Dataset:

In order to create a large mask dataset, the following algorithm is implemented. This algorithm is inspired by Liu et al. [6]

The process of this method is:

- Take two consecutive frames from a video, t and (t+1) as input
- Convert to grayscale
- Compute absolute difference
- Apply threshold
- Morphological operations
- Apply adaptive threshold
- Apply dilation

- Negate the image
- Resize

The flow diagram of the process is described in figure 4.1

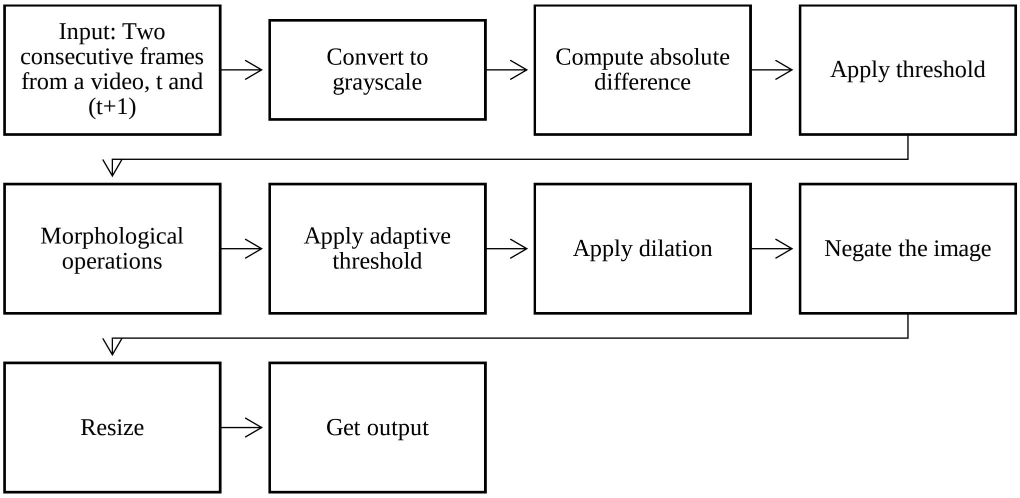

Figure 4.1: Process of mask synthetization using occlusion/dis-occlusion.

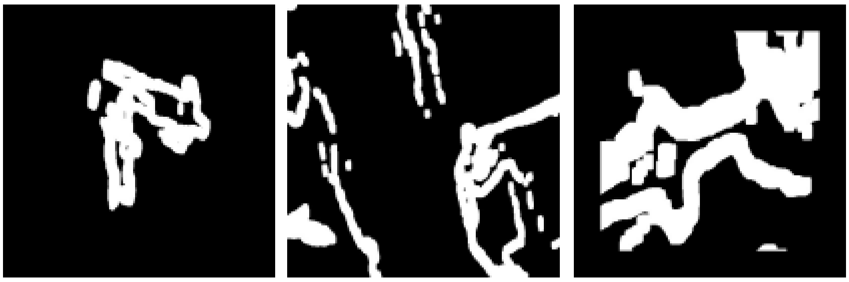

Figure 4.2: Generated masks using occlusion/dis-occlusion method.

This crafted dataset plays a vital role in offering a wide and varied platform for testing, enabling to thoroughly evaluate the proposed method. With its ability to handle various challenges, this dataset greatly enhances the comprehensive assessment and validation of the approach. As it was previously decided to target this thesis for natural images, Places2 [17]dataset has been chosen to be superimposed with the mask dataset. It is considered a good choice for image inpainting, due to its vast and diverse collection of more than 10Million images, spanning over 400 unique categories. This diversity can ensure that the model trained on it will be able to handle a wide range of inpainting scenario. Following are some example scenario from Places2 dataset. The images of this dataset are of 256x256 size, hence the masks curated using occlusion/dis-occlusion method are also of size 256x256.

Figure 4.3: Example of images from places2, subclass: lakes, forest, monument, stage

Due to the limitation of hardware available to train such an enormous dataset of Places2, which consists of a subset of 55,116 images were selected randomly from the dataset to be superimposed with the 55,116 masks. During training, a mask and an image from the Place2 dataset were randomly chosen to be superimposed with, and fed to the generator for training purpose. The generator would then try to generate the masked region, and continue with the training as described in section 3.2.

Following is an example of a ground truth, synthesized with the mask to create a deformed image, which later will be inpainted.

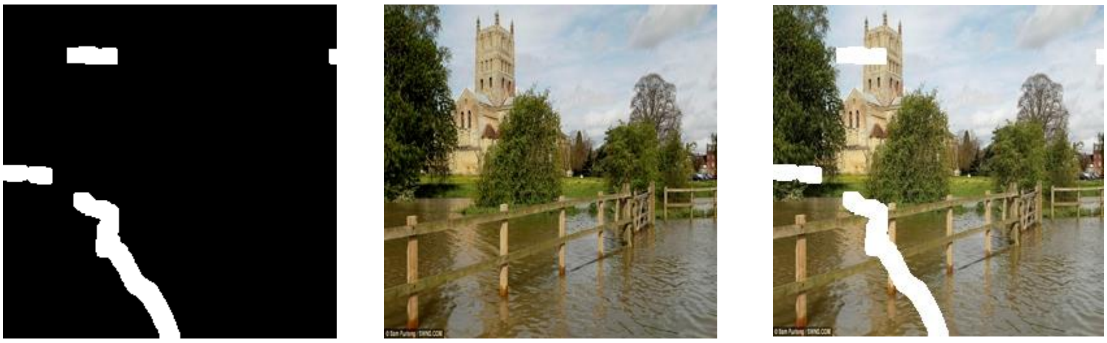

Figure 4.4: Synthesized mask, Ground truth, Deformed image

4.4 Result.

The result of this thesis can be evaluated in 2 ways:

I) Quantitative Result

II) Qualitative Result

4.4.1 Quantitative Result:

To evaluate the result of the thesis, 4 matrices has been chosen. L1 loss, L2 loss, PSNR, and SSIM. Here’s a brief introduction about the loss functions:

- **L1 Loss (Mean Absolute Error)**:

  L1 loss calculates the absolute differences between the predicted and ground truth values, then averages them , this loss is widely used in regression tasks, including image-to-image translation, where it penalizes large errors linearly. L1 loss is robust to outliers and encourages solutions with sparse error distributions.

$$L1(I,J) = \frac{1}{N}\sum_{i=1}^{N} |I_i - J_i|$$

- **L2 Loss (Mean Squared Error)**:

  L2 loss calculates the squared differences between the predicted and ground truth values, then averages them. It is commonly used in optimization problems and regression tasks, including image reconstruction and inpainting. L2 loss penalizes large errors more severely than L1 loss, which can lead to smoother solutions. But, L2 loss is sensitive to outliers and may prioritize reducing the magnitude of errors over preserving details.

$$L1(I,J) = \frac{1}{N}\sum_{i=1}^{N}(I_i - J_i)^2$$

- **PSNR (Peak Signal-to-Noise Ratio)**:

  PSNR measures the quality of a reconstructed image compared to the original image by calculating the ratio of the maximum possible pixel value to the mean squared error between the two images.PSNR is a popular metric for evaluating image quality in compression, reconstruction, and inpainting tasks.PSNR provides a simple and intuitive measure of image fidelity, where higher values indicate better reconstruction quality.PSNR does not always correlate well with human perception of image quality, particularly in scenarios where structural changes occur.

$$PSNR(I,J) = 20 \cdot \log_{10} \frac{MAX_I}{\sqrt{MSE(I,J)}}$$

- **SSIM (Structural Similarity Index)**:

  SSIM quantifies the similarity between two images based on luminance, contrast, and structure similarities, producing a value between -1 and 1. SSIM is widely used as a perceptual metric for evaluating image quality in tasks such as compression, denoising, and super-resolution. SSIM takes into account both global and local image features, making it more robust to structural changes and perceptually meaningful. SSIM requires more computational resources compared to PSNR and may not always capture perceptual differences accurately, especially in complex scenes.

$$SSIM(I,J) = \frac{\left(2\mu_I\mu_j + c_1\right).\left(2\sigma_{IJ} + c_2\right)}{\left(\mu_I^2 + \mu_J^2 + c_1\right).\left(\sigma_I^2 + \sigma_J^2 + c_2\right)}$$

Here,

- μ_I, μ_j are the the average intensities of image I,J,
- σ_I^2, σ_J^2 are their variances
- σ_{IJ} is the co-variance of the image
- $c_1 =(k_1 L)^2$, and $c_2 =(k_2 L)^2$ are constants to stabilize the division with weak denominator
- L is the dynamic range of the pixel-values, which is 255 for 8 bit images
- $k_1 = 0.01$ and $k_2 =0.03$ by default

- **IS (Inception Score):**

  Inception score evaluates the quality and diversity of the generated images by a model, using the inception network to classify images into pre-defined categories. High score indicate both realistic images and diversity. However, since it relies on the Inception model, it does not account for the match between generated images and input conditions, that can misinterpret the model's capability. Hence inception score may not be a good metric to judge inpainting model as described in [8]. Following is the equation of inception score, interested readers can read more at [18], [19]

$$\mathrm{D}_{\mathrm{KL}}\,(\mathrm{p(y|x)||p(y)} \;=\; \mathrm{p(y|x)log}\,\frac{\mathrm{p(y|x)}}{\mathrm{p(y)}}$$

$$\mathrm{IS} \;=\; \exp(\mathrm{E}_{\mathrm{x\sim pg}}\,\mathrm{D}_{\mathrm{KL}}\,(\mathrm{p(y|x)||p(y)})$$

The extracted generator is then being tested on an dataset of 500 images, and following is the result of the test:

The model is being tested 500 images on 4 conditions:

1) Using irregular mask with weight map at the border.
2) Using irregular mask without weight map at the border.
3) Using rectangular mask with weight map at the border.
4) Using rectangular mask without weight map at the border.

The results of the 4 conditions are tabulated in the following:

Table 4.1: Quantitative measurement of the aforementioned conditions on 500 images

Condition	L1 loss	L2 loss	PSNR	SSIM	IS
1)	7.8%	3.4%	16.6	0.60	1.081
2)	5.1%	2.1%	19.6	0.64	1.083
3)	1.3%	0.6%	28.98	0.94	1.088
4)	1.8%	0.8%	24.64	0.93	1.101

The following table shows a comparison between some of the image inpainting models and the result of the thesis. The comparison is done on the validation set of Places2 dataset A comparison is shown on both rectangular and free-form masks.

Table 4.2: Results of mean l1 error and mean l2 error on validation images of Places2 (irregular mask) [7]

Method	L1 loss	L2 loss
PatchMatch[[20]	11.3	2.4
Global & Local [21]	21.6	7.1
Contextual Attention[8]	17.2	4.7
Partial Conv[6]	10.4	1.9
Gated Conv[7]	9.1	1.6
Thesis	1.8	0.7

Table 4.3: Results of mean l1 error and mean l2 error on validation images of Places2 (rectangular mask)[7]

Method	L1 loss	L2 loss
PatchMatch[[20]	11.3	2.4
Global & Local [21]	21.6	7.1
Contextual Attention[8]	17.2	4.7
Partial Conv[6]	10.4	1.9
Gated Conv[7]	9.1	1.6
Thesis	1.6	0.6

4.4.2 Qualititive Results

The model has been trained on Places2 dataset and it is also tested on a few example from the same dataset. Some example of the model has been shown in the following. The PSNR and SSIM are 30.14 and 0.96 for the first image, and 23.5 and 0.95 for the second image. When the value of PSNR and SSIM are high, it indicates the output is closer to the original images.

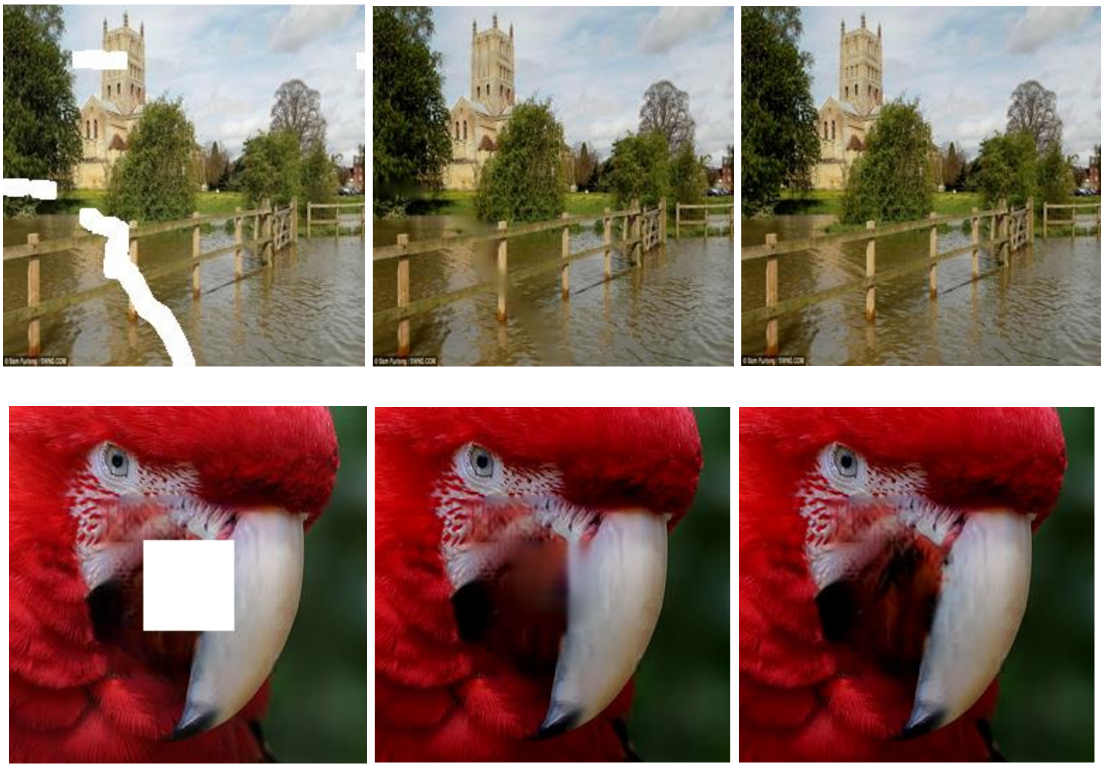

Figure 4.5: From left to right: masked image, inpainted image, ground truth

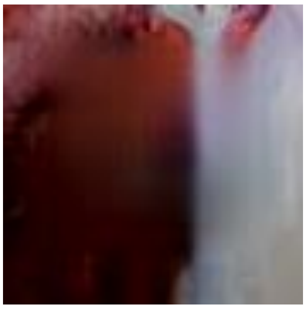

Figure 4.6: The inpainted region has been enlarged for better understanding.

Another use of image inpainting is object removal, a well-trained Image inpainting model can be used to remove object in an image, as much as being able to restore them, here are some example of the model being used for object removal. But since the inpainted object and the ground truth will have significant different in this particular case, the loss will be more noticeable despite the quality of the image. For example, in the first example of 4.6, the PSNR of the image is 15.98 and the SSIM is 0.92, the PSNR is low because some object has been removed from the original image, so the inpainted image is significantly different from the real image.

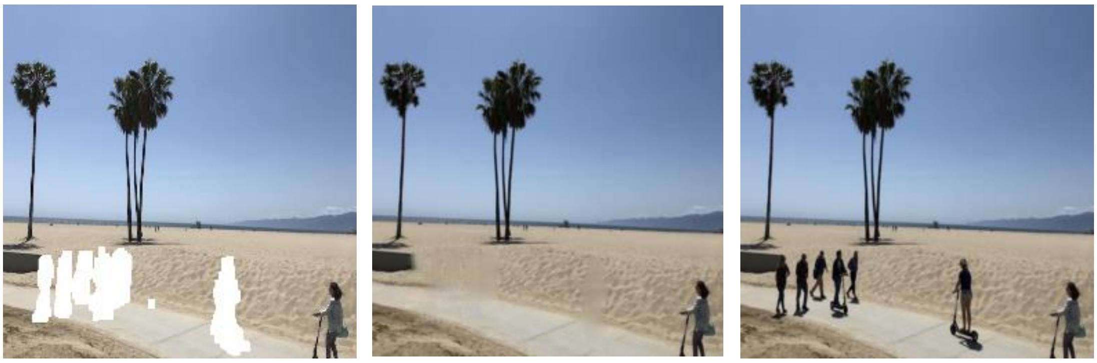

Figure 4.7: Masked image, inpainted output, ground truth in terms of object removal

Some results are not up to the mask, as it was expected. The model fails for some output. The result do not look realistic, rather it appears blurry. The following example depicts one such example, where the output of the model is not realistic. This needs much post-processing for the output to look as good.

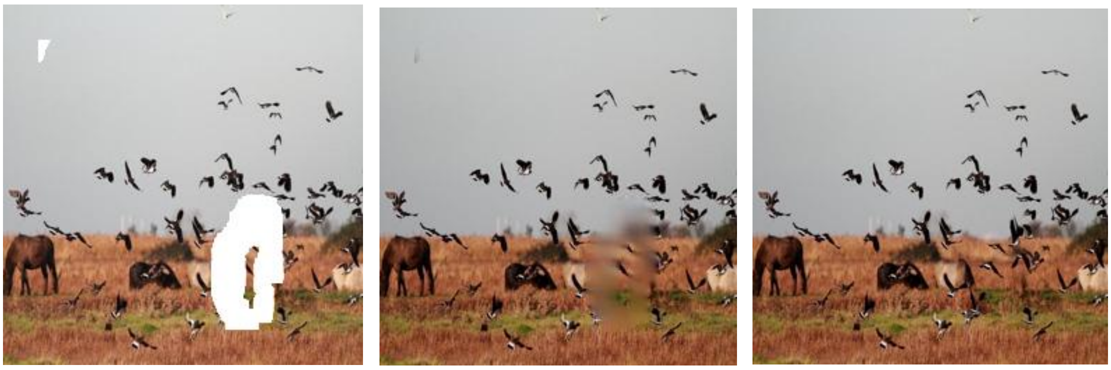

Figure 4.8: Masked image, inpainted output, ground truth

4.4.3 Analysis of Result

The result of the thesis has been compared with the state-of -the-art models. Most of the model outperforms the thesis, but the l1 loss is significantly lower than most of the state-of-the-art model. That is because, while calculating total loss, per pixel loss has been given more weight. This ensures the generated image appear closer to the real image. However, while doing so, the overall texture of the synthesized image has been overlooked which is why the inpainted regions looks blurred compared to the surrounding area for larger holes. Another possible result of this could be, this model has been trained only on 55,116 images, the test was conducted on 500 images, and the validation was performed on 200 images, Whereas, rest of the models have been trained on CelebA, Places2, DTD Texture, and other datasets along with the entire Places2 dataset. Due to the unavailability of hardware resources, it was not possible to train the model on such large datasets and had to opt for only a subset of one dataset. If the model was trained on same level, the model could perform better.

The thesis was conducted to be able to inpaint only irregularly shaped holes, but it can be seen that the model performs really well for geocentrically shaped holes as well. For geometrically shaped hole rectangle was chosen as the mask, maximum size of the hole has been one rectangle of size 128x128 or 2 rectangles of size 64x64. These parameters were chosen to compare the result with [8]

4.5 Objective Achieved

The thesis, after being tested on the test set and evaluated compared to the state-of-the-art models, the assessments can be made on the previously declared objectives of the thesis:

- The image inpainting generator is based on a Conditional GAN, as it was previously discussed. The target was to accomplish it by conditioning the generator on the binary mask to inpaint the targeted region. The method works, and the model can reconstruct the missing region using the mask as a condition.
- A study has been made to to check the effect of prioritizing the border region whilw inpainting, and it is observed that using a weight map to give more importance in the border region generates better output. The SSIM and PSNR values are significantly better.

- The target to inpaint larger region realistically has not been accomplished succcesfully, while the generator can reconstruc the region, it looks blurred, and not realistic.
- The thesis has been compared against the state-of-the-art models, and it can bee seen the model performsreally good for the geometrically shaped masks with a higher PSNR and SSIM.

4.6 Financial Analysis and Budget

This thesis project is conducted with careful consideration of financial aspects. The allocation of resources for data acquisition, experimentation, and computational support is thoughtfully managed within the available budget. To ensure efficiency, the utilization of open-source tools and institutional facilities has been priotized. While software requirements are met, the aim aim to maximize cost-effectiveness by utilizing free IDEs like Google Collab or Kaggle. Given the expense associated with acquiring extensive image datasets, the use of public datasets and is prioritized and their creators are duly acknowledged. Additionally, provisions are made to accommodate unforeseen expenses, ensuring the project's smooth progression without compromising quality. Table 4.4 gives an idea of the total budget allocation for conducting the thesis.

Table 4.4: Financial Budget Plan

Budget Category	Estimated Cost Range
Equipment and Software	10,000 - 30,000 BDT
Dataset	5,000 - 10,000 BDT
Travel and Conferences	2,000 - 5,000 BDT
Publication Fees	1,000 - 3,000 BDT
Miscellaneous	1,000 - 3,000 BDT
Total Estimated Budget Range	19,000 - 51,000 BDT

By adhering to a strategic financial plan, which includes a designated budget in case the current resources fall short of meeting criteria, this thesis aims to achieve optimal outcomes while maintaining financial responsibility.

4.7 Conclusion

In conclusion, the implementation phase has encompassed essential tasks such as dataset curation, technical setup, and methodology formulation. The implementation and result phase has been most important in the fine tuning of the model and the generator architecture. The goal has been to minimize the l1, l2 loss and increase the PSNR and SSIM of the generator. This evaluation matrices dictates the quality of the thesis.

CHAPTER V

Societal, Health, Environment, Safety, Ethical, Legal and Cultural Issues

5.1 Introduction

This report encompasses the socio-economic impact and the ethical considerations in the field of image inpainting. It highlights the accomplishments so far and sets the stage for the project's continued evolution.

5.2 Intellectual Property Consideration

As part of this thesis, careful attention is given to both the legal and ethical dimensions of utilizing datasets and models. In addition to ethical practices, all datasets and models are acquired and employed with the proper permissions and authorization from their respective authors or owners. This ensures that intellectual property rights and usage terms are respected. Additionally, the considerations extend to proper attribution and acknowledgment of the original authors' contributions. For example, here the Places2[22] dataset has been used, and credited. At the same time, the test-mask for the testing has been acquired from [6]and been cited respectively. Not only does this uphold scholarly integrity, but it also fosters a collaborative and respectful academic environment. By doing so, this thesis ensures responsible data usage, respects creators' rights, and promotes transparency in research practices.

5.3 Ethical Consideration

There are several ethical concerns with the use of image inpainting technology. Concerns are raised by the advanced inpainting technology's ability to edit images with high fidelity. Given that generative networks have the capacity to produce harmful and deceptive content—such as deepfakes—it is crucial that moral guidelines and standards be established to prevent any improper use of this technology. In doing the thesis, caution was taken to ensure that no privacy was violated by the data used to train the model. Furthermore, it is now crucial to establish moral standards for usage.

5.4 Safety Consideration

The thesis required the model to be trained on a large dataset, and originality of the dataset was a concern throughout the whole process. The data had been preprocessed and post processed, but it had been made sure of the fact that, no original information carried in the data or image do not get altered in the process. This ensures the security of the image inpainting model. It only generates the distorted regions, keeping the original photo intact. It can be said this model carries the original information as well as generates new information.

5.5 Legal Consideration

This thesis's legal considerations center on a few issues, such as liability, intellectual property, and copyright. To begin with, the dataset that was used for testing and training was made available to the public and did not contain any copyrighted material. No personal or private information was used for this thesis's research without the owners' permission. Furthermore, any generative network's development carries the risk of producing deceptive images produced by AI. Since users are ultimately responsible for these images, the application should be closely watched.

5.6 Impact of the Project on Social, Health, and Cultural Issues

This thesis, and by extension the image inpainting technology using any generative network, impacts the surrounding social life, along with the health sector and cultural sector. This technology can be used in forensic science to conduct in-depth investigations on partially occluded images, though more research is necessary. It can assist in reconstructing the damaged areas of MRI or CT scan images. Image inpainting plays an socially valuable role by facilitating the restoration of old photos or intimate memories, protecting cultural heritage and history. This technology can also generate, and complete partially cropped or hidden human faces when trained over enough data. Thus, it is possible to use this to preserve cultures and customs. An important application of image inpainting with cGAN is restoring historical predated images that have been distorted over time to their original form with ease and without sacrificing their uniqueness.

5.7 Impact of Project on the Environment and Sustainability

Image inpainting technology has a profound impact on the environment and sustainability aspects. Although the number of resources required to train these models requires powerful GPU or specialized hardware for extended periods, consuming electricity, the technology has promising potential in conserving digital restoration. This process has reduced the need for physical and chemical materials traditionally used in the field restoration. Also, it's possible to compress images more efficiently for storage and transmission. For example, transmitting partially complete images and then restoring them to their original state can save bandwidth and reduce energy consumption, making it a potentially sustainable technology that can be further investigated.

Socio-Economic Impact:

1. **Enhanced Visual Content:** The development of advanced image inpainting techniques contributes to improved visual content across various industries, such as media, advertising, and entertainment. By seamlessly restoring and enhancing images, the thesis supports the creation of more engaging and appealing visual materials.

2. **Medical Imaging and Forensics:** The application of image inpainting in medical imaging aids in the reconstruction of missing or corrupted medical data, potentially leading to better diagnostic accuracy. Additionally, in the field of forensics, the restoration of obscured details in images can provide crucial evidence for investigative purposes.

3. **Cultural Heritage Preservation:** Image inpainting holds significance in art restoration and cultural heritage preservation. By accurately reconstructing missing

portions of historical artworks or artifacts, the thesis contributes to the conservation of our cultural heritage for future generations.

Sustainability:

1. **Efficiency and Resource Optimization:** Improved image inpainting methods streamline post-production processes in industries like photography and film, saving time and resources. This efficiency contributes to sustainable production practices and reduced energy consumption.
2. **Waste Reduction:** Image inpainting techniques can aid in salvageable image restoration, minimizing the need to recreate images from scratch. This waste reduction aligns with sustainable practices by conserving resources and reducing environmental impact.
3. **Longevity of Visual Content:** By facilitating the restoration of damaged or deteriorated visual content, the thesis promotes the reuse of existing materials, reducing the demand for new content creation. This sustainable approach aligns with efforts to minimize the carbon footprint of digital content generation.
4. **Cross-Disciplinary Collaboration:** The knowledge gained from the thesis can be shared across disciplines, fostering interdisciplinary collaborations. This interdisciplinary approach encourages the exchange of ideas and solutions, leading to more sustainable innovation.

The thesis on image inpainting brings about socio-economic benefits by enhancing visual content quality and aiding various fields. Its sustainability lies in promoting resource efficiency, waste reduction, and interdisciplinary collaboration. By addressing both short-term impact and long-term sustainability, the thesis contributes to a more robust and responsible digital ecosystem.

5.8 Conclusion

The goal of any research project is to advance humanity and society, and the thesis seeks to accomplish this as well. While every technology can be misused by the wrong hand, it is necessary to set up a legal framework and ethical code of conduct for technologies like this to prevent them from being used with a malicious intent. Instead, attention should be placed on the beneficial uses of image inpainting, which are found in the pursuit of durability and sustainability.

CHAPTER VI

Addressing Complex Engineering Problem Associated with Image Inpainting

6.1 Introduction

Image inpainting is one of the trending computer vision problems currently. In recent years, significant advancements have been made in image inpainting, the credit goes to the fast adaptation of deep learning techniques and the availability of the large scale dataset. Despite the advancements, there are still a lot of scopes of improvement left in the problem. This chapter aims to discuss some of the problems associated with the field of image inpainting, as well as the complex activities associated with the research.

6.2 Complex Engineering Problems Associated with Image Inpainting

This section discusses the problems associated with image inpainting in general, for example:

- Most of the existing datasets are too homogeneous or small or too synthetic to capture the complexity and variability of the real-world images.
- To ensure the structural and semantic consistency and coherence of an inpainted region with the rest of the image is a challenge. This means the regions should look natural and realistic and match the surrounding area as well.
- Being able to control the inpainting, that is having gained user-control is difficult, because the model can only take input from the user, there is not much scope for variability and customization in this field yet[23]
- Computational efficiency and scalability is still one of the largest concern in this field. A generative model usually takes up a lot of time to be trained, even in an

advanced computer with multiple GPU or computational unit. Yet, the number of datasets cannot be reduced, because it is crucial the model is trained on enough data to produce a realistic image afterwards. So for a model to generate realistic output, it needs to be trained for days.

6.3 Complex Engineering Activities Associated with Image Inpainting

The process of training an image inpainting model using a generative adversarial network is a rigorous task that involves many steps, apart from designing the architecture or training. There are many parameters that needs to be considered and calculated based on the target dataset, or goal of the model. This includes many mathematical calculations, as well as pre-processing, and post processing of the image. Before, starting of with the model, it is required to decide upon the target of the model, hence prepare a suitable dataset. Any machine learning model requires a large number of data, which is not often available in many cases, for example, in the medical domain. So, it is necessary to perform proper data augmentation to increase the number of images in a dataset, and also to make sure the augmentation is appropriate for robust and generalized model.

Then it is necessary to study and explore various architectures, that maybe suitable for the goal. This requires an extensive study of Convolutional Neural Network, Generative Adversarial Network, various Image Processing concepts, and Machine Learning concepts to help design the inside and outside of image inpainting.

To train the model accurately, the most important part of the training is loss calculation. Defining appropriate loss function is crucial to the training because, the back propagation to adjust the weight of the model is based on the loss values. [24]

Once the training is complete, it is necessary to check and test the model against state-of-the-art methods and evaluate using the evaluation matrices. This tells how accurate a model is. There are different evaluation matrices for different purpose, it is necessary to identify the chose the right one for the right task. It will be of poor judgement to choose L1 or L2 loss for a task where the quality of the image is more important that pixel loss, at the same time, when accuracy is more important, it is necessary to use MSE or MAE loss instead of inception score.

6.4 Conclusion

In conclusion, image inpainting is undoubtedly a fascinating yet challenging domain that integrates computer vision, machine learning and image processing. In this chapter the current problems and complex activities that are associated with the field has been discussed. Despite the progress, there still lacks methods to reconstruct larger regions, or reconstruction of region with lesser available dataset. Ensuring temporal and spatial consistency in these fields has remained a problem for a long time. Balancing texture synthesis and maintaining structural coherence at the same time is a important factor that should be studied more in the future. Also, the complex pre-processing and post-processing activities should also be looked into for the improvement of the result.

CHAPTER VII

Conclusion

The report encompasses the journey, progress, methodology and the considerations in the field of image inpainting. It also highlights the accomplishments and the shortcomings, which set the stage for the next stage of the research.

7.1 Conclusion

Image inpainting is one of the advanced and important field in computer vision. The application of this lies from medical imaging to augmented reality. As it can be seen, the GAN based models are more diverse, and accurate than traditional Convolutional Neural Network based models. The images generated from Generative Adversarial Networks happen to be more realistic and semantically coherent than the ones generated from traditional inpainting methods. The thesis can be divided into two broad category, one : when a weight map is used to specify a border region, the other one is used without weight map. It can be seen, when image inpainting is done using a weight map, the generator performs better.

7.2 Limitations

The thesis was conducted within a limited time, and the resources were limited as well. It was not possible to train the model on multiple datasets. Infact, training the model on the Places2 dataset[17] , which consists of more than 2.1M images were time consuming, so the model was trained on a subset of 55000 images from places2 dataset. Had the model been trained on the entire dataset, it would have performed better. Most other literatures that have been reviewed for the experimentation of the thesis has used more than one datasets including CelebA dataset[25], DTD textures [26]dataset, etc. The hardware limitation played a major factor in this regard. Also due to the same reason. The generator architecture consists of only 5 layers. A deeper network on a deeper dataset should perform better. Another limitation for the model is that, this model has been trained on image size 256x256, which is relatively small.

7.3 Future Works

Looking ahead, the future work entails further refinement of the model's architecture and hyperparameters for enhanced performance, experimenting with novel loss functions to improve realism, conducting comprehensive evaluations on diverse datasets to understand the strength and limitations, incorporating user feedback to develop more intuitive tools, explore real time capabilities, considering transfer learning to expedite training, and upholding ethical practices in data usage. These steps collectively aim to advance the models effectiveness, versatility and ethical integrity.

In conclusion, this report outlines the foundational steps taken in the pursuit of advancing image inpainting techniques. With a focus on technical progress and ethical considerations, the groundwork has been laid for the development and evaluation of the model. As the thesis concludes, hopefully this can be helpful in the advancement of the research of image inpainting to come.

REFERENCES

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image Inpainting”.

[2] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and Locally Consistent Image Completion,” *ACM Trans. Graph*, vol. 36, 2017, doi: 10.1145/3072959.3073659.

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

[4] S. Yan and X. Zhang, “PCNet: partial convolution attention mechanism for image inpainting,” *International Journal of Computers and Applications*, vol. 44, no. 8, pp. 738–745, Aug. 2022, doi: 10.1080/1206212X.2021.1909280.

[5] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 5967–5976, Nov. 2017, doi: 10.1109/CVPR.2017.632.

[6] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, “Image Inpainting for Irregular Holes Using Partial Convolutions,” Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.07723

[7] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, “Free-Form Image Inpainting with Gated Convolution,” Jun. 2018, [Online]. Available: http://arxiv.org/abs/1806.03589

[8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative Image Inpainting with Contextual Attention,” Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.07892

[9] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” Nov. 2014, Accessed: Feb. 14, 2024. [Online]. Available: http://arxiv.org/abs/1411.1784

[10] V. Nair and G. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair,” in *Proceedings of ICML*, Feb. 2010, pp. 807–814.

[11] G. Bingham and R. Miikkulainen, “Discovering Parametric Activation Functions.” Feb. 2020.

[12] Y. Lecun, L. Bottou, G. Orr, and K.-R. Müller, “Efficient BackProp,” Feb. 2000.

[13] I. Drori, D. Cohen-Or, and H. Yeshurun, “Fragment-based image completion,” *ACM Trans Graph*, pp. 303–312, 2003, doi: 10.1145/1201775.882267.

[14] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep Image Prior,” Nov. 2017, doi: 10.1007/s11263-020-01303-4.

[15] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, “Semantic Image Inpainting with Deep Generative Models,” Jul. 2016, [Online]. Available: http://arxiv.org/abs/1607.07539

[16] R. Liu *et al.*, “SCCGAN: Style and Characters Inpainting Based on CGAN”, doi: 10.1007/s11036-020-01717-x/Published.

[17] "Places-2_MIT_Dataset." Accessed: Feb. 15, 2024. [Online]. Available: https://www.kaggle.com/datasets/nickj26/places2-mit-dataset

[18] PyTorch-Ignite Contributors, "INCEPTIONSCORE."

[19] S. Barratt and R. Sharma, "A Note on the Inception Score", Accessed: Feb. 20, 2024. [Online]. Available: https://github.com/

[20] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, Jul. 2009, doi: 10.1145/1531326.1531330.

[21] "Iizuka, Satoshi, Edgar Simo-Serra, and Hiroshi Ishikawa. 'Globally and Locally Consistent Image Completion.' ACM Transactions on Graphics (TOG) 36.4 (2017): 107. - Google Search." Accessed: Feb. 15, 2024. [Online].

[22] 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, doi: 10.1109/TPAMI.2017.2723009.

[23] "What are some of the current challenges and limitations of image inpainting techniques?" Accessed: Feb. 15, 2024. [Online]. Available: https://www.linkedin.com/advice/0/what-some-current-challenges-limitations-image

[24] "Globally and Locally Consistent Image Completion(이미지 복구)." Accessed: Feb. 15, 2024. [Online]. Available: https://jayhey.github.io/deep%20learning/2018/01/05/image_completion/

[25] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild *," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, 2015, Accessed: Feb. 20, 2024. [Online]. Available: http://personal.ie.cuhk.edu.hk/

[26] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing Textures in the Wild".