Enhancing Image Reconstruction with Prediction model using Deep Convolutional GANs

MSc Research Project
Data Analytics

Sahil Bhujbal
Student ID: x18183395

School of Computing
National College of Ireland

Supervisor: Prof. Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sahil Rajesh Bhujbal		
Student ID:	x18183395		
Programme:	MSc in Data Analytics	**Year:**	2019-2020
Module:	MSc Research Project		
Supervisor:	Prof. Vladimir Milosavljevic		
Submission Due Date:	28th September, 2020		
Project Title:	Enhancing Image Reconstruction with Prediction model using Deep Convolutional GANs		
Word Count:	7962 **Page Count:** 22		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ....

Date: ....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Image Reconstruction with Prediction model using Deep Convolutional GANs

Sahil Bhujbal
x18183395

Abstract

Image Inpainting has gained a lot of popularity within the applications of Computer Vision and Image Processing due to the ability of generating, restoring and modifying images and videos. Current Image Inpainting methods which are incorporated with Generative Adversarial Networks (GANs) can generate feasible inpainting results. However, previous methods are susceptible to fall into the Mode-Collapse situation due to inadequate training. This research aims to avoid the situation of Mode-Collapse during the training phase by proposing a novel Image Prediction model along with the Deep Convolutional GAN (DCGAN) model. Upon evaluation of the model, results can be backpropagated to the DCGAN model to enhance training, prediction and inpainting of images. Additionally, three datasets are selected for performing experiments based on three different scenarios related to the complexity of images. All images are evaluated in the prediction phase with the help of Inception Score metric and model losses in the training and inpainting phase. Experimental results provide good insights regarding image prediction, thus proving beneficial in enhancing the reconstruction technique.

Index Terms: **Image Inpainting, Deep Convolutional Generative Adversarial Network (DCGAN), Image Prediction, Inception Score**

1 Introduction

1.1 Background Overview

The use of Generative Adversarial Networks has taken an exponential step in the field of both Machine Learning as well as Computer Vision. Since the introduction of GANs by (Goodfellow et al., 2014) in 2014, there have been many amendments in the architecture and the algorithm of the original model by various researchers for obtaining better results, qualitatively as well as quantitatively. Also, the implementation of GANs have been extended to various technical applications in image and video processing such as Classification, Detection, Segmentation, Reconstruction, etc. With the invention of Deep Convolutional Generative Adversarial Networks (DCGAN) in 2016 by (Radford, Metz and Chintala, 2016), it boosted the usage of the Generative models amongst many researchers and data engineers due to its robust nature and qualitative representations. The use of DCGANs in various Computer Vision applications have led the way to explore the usage of the Generative models.

1.2 Motivation

Image Inpainting is one of the applications of Computer Vision which aims in filling the missing or corrupted regions from the images. The technique takes surrounding pixels into consideration and tries to complete the image. There have been different adaptations in performing Image Inpainting using Auto-Encoders, Partial Convolutions, Exemplar-based Methods, Photoshop applications, and Generative Networks. All of them have achieved meaningful results in the objective they were trying to achieve. However, there are cases where all these methods fail due to Mode-Collapse situation at times and fails to deliver a realistic output image. Mode-Collapse is a situation where the Generator Network fails to generate realistic images and the loss function spikes to a value which cannot be reduced. Also, when Generative Networks are in used for Inpainting, Image Prediction is not carried out which serves as a bridge between Training and Inpainting. The methods have really focused on enhancing either their Training models or the Inpainting models. These Inpainting models have been previously focused on denoising images using Generative Networks and patch filling using other methods. Due to the absence of the Image Prediction, Inpainting models generate higher losses for the reconstructed images.

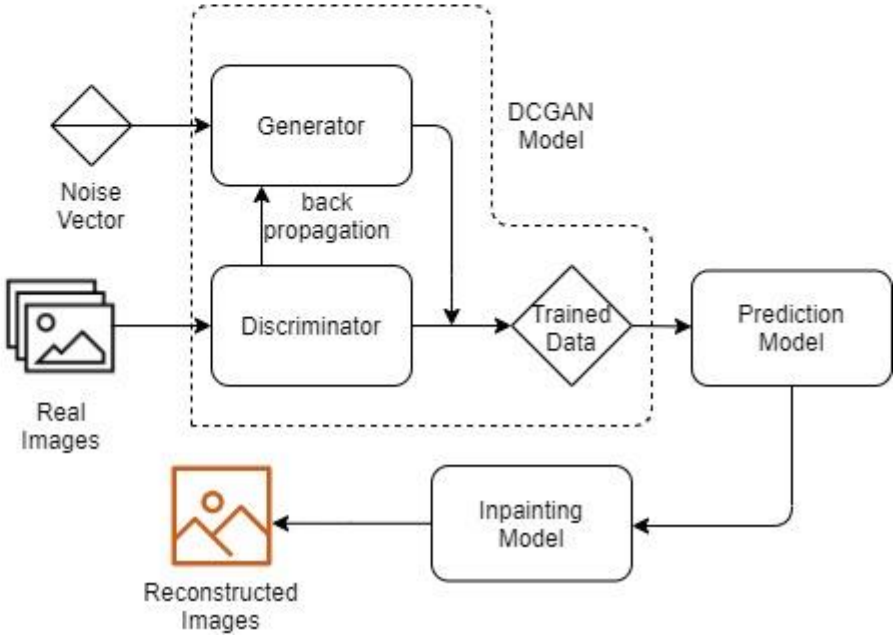


Figure 1: Basic Model Implementation[1]

1.3 Research Question

"How can a Deep Convolutional GAN model incorporated with Image Prediction, enhance the Image Inpainting application?"

1.4 Research Objective

In this proposed research, the gap is aimed to fill with the introduction of Image Prediction and passing it to the Inpainting model to avoid the Mode-Collapse scenario and reduce reconstruction losses to minimal for obtaining semantic results of corrupt images. Also, the dataset selected for the research is the Celeb-A Faces dataset in order to try and generate face images after applying mask to the input and considering the Global Pandemic situation as a future application. Experiment on Celeb-A Faces dataset is compared with experiments performed on two other datasets taking complexity of the image into consideration, which is explained briefly in the Evaluation section. The proposed research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology and makes use of the DCGAN model for training the data with changes in hyperparameters for generating semantic images. Upon completion of training, instead of performing Image Inpainting, Image Prediction task

[1] Research flowchart is self-drawn after understanding the process.

has been carried out with the help of the pre-trained DCGAN model. It resulted into critical evaluation of the DCGAN model identifying the qualitative values of generated images using Evaluation Metrics. The third step covers validating the trained model by performing Image Inpainting based on Image Prediction, taking two loss terms into consideration and resulting into representative Image Generation.

The structure of the report is as follows. Section 2 covers an in-depth literature review about the preceding works on different GANs and various Image Inpainting methods. Section 3 covers the proposed methodology for the research followed by Design Specification and Implementation in Section 4. Critical model evaluation is performed in Section 5 and Section 6 discusses the Conclusion, limitations, and future prospect regarding the proposed model.

2 Related Work

There have been various Inpainting implementations in the past in order to acquire realistic reconstruction results. A detailed literature review has been carried out on those implementations and, various types of GANs have been studied in order to select and modify the best suitable GAN model for implementing the Prediction and the Inpainting model.

2.1 Overview of Different GAN methodologies

The first ever Generative Adversarial Network was proposed by (Goodfellow *et al.*, 2014) or can be called as the Vanilla GAN model explaining all the aspects of the model in deep. They defined two neural networks, a Generator network and a Discriminator network and trained them simultaneously but adversarial to decide if the input is from a generated sample or the main data. They figured out that training the GAN was a tremendously difficult task and had to face various issues like over-fitting, mode-collapse, etc. Thus, this model led many researchers to tune various hyperparameters and start the journey of developing different application specific GANs.

(Radford, Metz and Chintala, 2016) have introduced a new class of Convolutional Neural Network models known as Deep Convolution in GANs to increase the efficiency of Unsupervised Learning in the field of Computer Vision. They implemented this updated model on three different public datasets for training and used it as a pre-trained model for performing classification on the CIFAR-10 dataset. This was really a revolution in the field of Generative Networks as the model easily outperformed various preceding techniques and it made a path of future data engineers to explore the model and implement in various Computer Vision applications.

A unique idea was proposed by (Huang *et al.*, 2017) of implementing a Stacked GANs in a top-down manner in order to take multiple noise input vectors into consideration and generate conditioned low-level representation of images. With the employment of two losses, conditional loss and an entropy loss, they train their model stack-by-stack before performing complete model training. They evaluated their methods with the nearest neighbor algorithm as well as using various evaluation metrics to find the best possible result and proved to be more reliable than a single GAN model.

A new upgraded GAN architecture was proposed by (Zhang *et al.*, 2019) known as the Self-Attention GAN (SAGAN) which takes all the feature points into consideration while generating details for images, thus providing high-resolution images as output. They have implemented the SoftMax function in their self-attention module and performed training on the benchmark ImageNet dataset. On comparison with previous state-of-the-art methods, their research provided a better Inception Score after evaluation.

In order to tackle the problem of vanishing gradients from the Discriminator network due to loss functions, authors have proposed a method of modifying the GAN model which generates a new least square loss function (Mao *et al.*, 2017). Their model is a combination of VGG and DCGAN, and the hidden layers are selected and mapped accordingly. They implemented the training on a dataset of handwritten Chinese characters and were able to produce readable output images. They aim to extend their research by implementing the model on more complex datasets and achieve similar results.

In this research (Anirudh *et al.*, 2019), authors proposed a technique in the medical field called as Computed Tomography (CT) reconstruction which is a procedure of reconstructing the object structure from sequences of x-ray images. This research makes use of a pre-trained robust GAN model for resolving issues of inverse problems which uses a method known as corruption mimicking for improving the qualitative values of image reconstruction. They train the DCGAN model on two public datasets and adjusts the hyperparameters as per corruption mimicking technique to obtain very good results of reconstruction. They have compared their results with previous state-of-the-art methods and outperformed them.

A unique model was proposed by (Liu and Tuzel, 2016) which combined two different GANs for implementing joint distribution on multi-domain images. They coupled a pair of GANs by tying weights in the Generator and the Discriminator networks to achieve learning of joint distribution. Model was also implemented on Celeb-A faces and MNIST datasets for checking multi-modality.

A different Generator architecture was proposed by (Karras, Laine and Aila, 2019) where they replaced the traditional Generator architecture of GAN with a style transfer-based architecture which consists of two different networks, namely Mapping network and Synthesis network. Their Generator network outperformed the traditional Generator network in all fields, and they were able to implement their state-of-the-art method using GAN.

(Yang *et al.*, 2019) proposed a technique for Conditional GANs to tackle the issue of mode-collapse in the Generator network. They regularized the Generator network explicitly in order to deliver varied outputs depending on latent codes. They compared their technique with existing conditional GANs and achieved great qualitative as well as quantitative results and completely avoided the mode-collapse problem faced in GANs.

By taking two GAN applications into consideration, (Salimans *et al.*, 2016) proposed various novel features and training measures to implement on GANs. First application covers the semi-supervised learning and the second covers Image Completion. As mentioned, they implemented various techniques to improve the performance of GAN like, Feature matching, Virtual batch normalization, Minibatch Discrimination, Historical Averaging and one-sided Label smoothing and evaluated using Inception Score metrics.

2.2 GAN based Image Inpainting

To tackle the inpainting ambiguity in GAN models, (Li *et al.*, 2019) proposed a technique to resolve the issue of semantic consistency. They implemented an Inpainting network with the Generator network to fill the mask in input images. They updated the Loss functions from the original GAN implementation and were able to provide better results overall.

In this paper (Demir and Unal, 2018), the authors have proposed a new method which combines the global GAN architecture with a patch-GAN approach and afterward parts ways to create two antagonistic misfortunes that feed the generator so as to catch both neighborhood coherence of picture surface and worldwide highlights in pictures. Generator comprised of a ResNet architecture. However, since the mask patches were of a fix size and focused in the center of images, the method was vulnerable to errors at the edges.

This paper focuses on designing a Pluralistic Image Inpainting model with the help of GAN in order to generate multiple images for a single input image with reasonable content (Cai

and Wei, 2020). This is achieved with the help of a style extractor which is used to feature extraction from images of ground truth. The pluralistic model clearly exceeds them in both qualitative as well as quantitative comparisons.
In order to reduce the losses in the Generative networks, (Xu, Liu and Xiong, 2018) proposed an idea of edge guided GANs for Image inpainting to fit in the edges inside the generative network. They modified their generative network by providing a down-sample from the Input edge along with the input vector.
This paper (Nazeri *et al.*, 2019) proposes an adversarial model for Image inpainting with two stages, EdgeConnect which is basically an edge generator and an Image completion network. The EdgeConnect network tries to generate the missing sections of the images with the help of edges and the Image completion network fills those missing sections with the results obtained from the edge network.
A novel idea was proposed by (Hui *et al.*, 2020) in order to tackle the problem of incomplete images with losses in pixels. The model included a refined Generative network, a one-phase model that uses thick blends of enlarged convolutions to get bigger and increasingly viable open fields. Thus, it was easier to fill large and inconsistent holes in images with help of a regression loss and geometrical alignment constraint which were introduced for enhancing semantic details of the image.
First ever implementation of Image inpainting using GANs was proposed by (H. Liu *et al.*, 2018). They were able to implement with the help of a Deep Convolutional GAN model with two losses, Context loss and Gradient Loss. The model was applied on the Celeb-A dataset and their exploratory outcomes showed that the results for the reconstructed images can keep up the worldwide consistency of the structure and the clearness of the nearby surface.
The authors of (Vitoria, Sintes and Ballester, 2019) made use of the Wasserstein GANs as they trained the network to learn the data space. Also, with the use of Wasserstein GAN, they were able to update and create new Generator and Discriminator architectures for training. They were able to combine the learned information with a new loss function for carrying out the Image inpainting task and minimizing the function provided help in filling the corrupted data neatly. This technique also takes the Contextual and the Perceptual information into consideration while performing Image inpainting.
To tackle the inconsistency of regenerating high-resolution images, (Yeh *et al.*, 2017) proposed an idea based on generating pixels by conditioning the input data. They made use of a pre-trained Generator model and along with two losses, they aimed in reconstructing corrupted images. They try to keep their generator updated with the data provided by updating the input value in order to find the closest value to the real images.
Inconsistency in Image inpainting results due to ineffective Convolutional Neural Networks and gap in Contextual information and Perceptual information led to a novel approach of using a contextual attention layer in GAN for Image inpainting (Yu *et al.*, 2018). This consists of two Dilated Convolutional Networks, a Course network and a Refinement network. They trained the first network with just the reconstruction loss and the second network with the reconstruction loss along with two Wasserstein GAN losses.

2.3 Other Image Inpainting Techniques

Authors proposed a method in light of joint enhancement of picture substance and surface requirements, which preserves logical structures as well as produces high-recurrence details by coordinating and adjusting patches with the most comparable mid-layer connections of a profound classification model (Yang *et al.*, 2017). They made use of the VGG-19 pre-trained model as their classification network and performed regression in the content network.
Authors have proposed two techniques for enhancing the algorithm used in Image inpainting using the Exemplar-based approach (Ishi, Singh and Agrawal, 2015). Firstly, in order to

reduce the complexity of searching, a gradient based searching is developed. Secondly, they have made used of patch propagation technique and have upgraded it with a distance-dependent criterion to maximize the accuracy.

An Exemplar-based Inpainting method was proposed by (Xu and Sun, 2010) to challenge the issues of image sparsity. In this model, two new concepts were introduced. To start with, they characterized a novel fix need in view of the scarceness of the patch's non-zero likenesses to its neighboring patches. They termed it as structure sparsity. With this, they were able to clearly distinguish between the structure and texture within images. Later, they made use of spare linear combination of exemplars to inpaint on the boundary of missing sections of the image. They termed it as patch sparse representation. Together, they establish the patch sparsity which is the novelty in the research.

This paper (G. Liu *et al.*, 2018) was published by Nvidia Corporation. The authors have made use of Irregular shaped masks to perform Image Inpainting and they have replaced GANs convolution layers with partial convolutional layers where the convolution is masked and renormalized to be conditioned on only valid pixels of Images. They implemented this with the help of VGG-16 pre-trained model and performed nearest neighbor up-sampling to generate an updated mask of the next layer.

In order to obtain better visual results with a fast runtime, (Hirvola *et al.*, 2016) proposed an Inpainting technique which showcased an ability to fill missing regions of images in an omni direction. This is an extension to the current exemplar-based image inpainting methods and takes the image contour point selection into consideration. With this method, they were able to perform image inpainting as well as object removal on public datasets.

Thus after carrying out an intense literature survey on different GAN methodologies and various Inpainting techniques, it has been observed that Image Prediction has not been implemented in any of the previous researchers and most of the methods which are using GANs are vulnerable to the Mode-Collapse situations. So, this research aims to take these limitations into consideration and build a Train-Predict-Inpaint model for reconstructing images.

3 Research Methodology

In order to tackle Data Mining related problems, it is essential to have a systematic method which follows a structured procedure to evaluate Machine Learning models and acquire sophisticated solutions for such problems. Cross-Industry Standard Process for Data Mining (CRISP-DM) Methodology just seems to be an ideal technique for this research project as it follows the structured procedure from understanding the business perspective of the research to its planned deployment in future prospect.

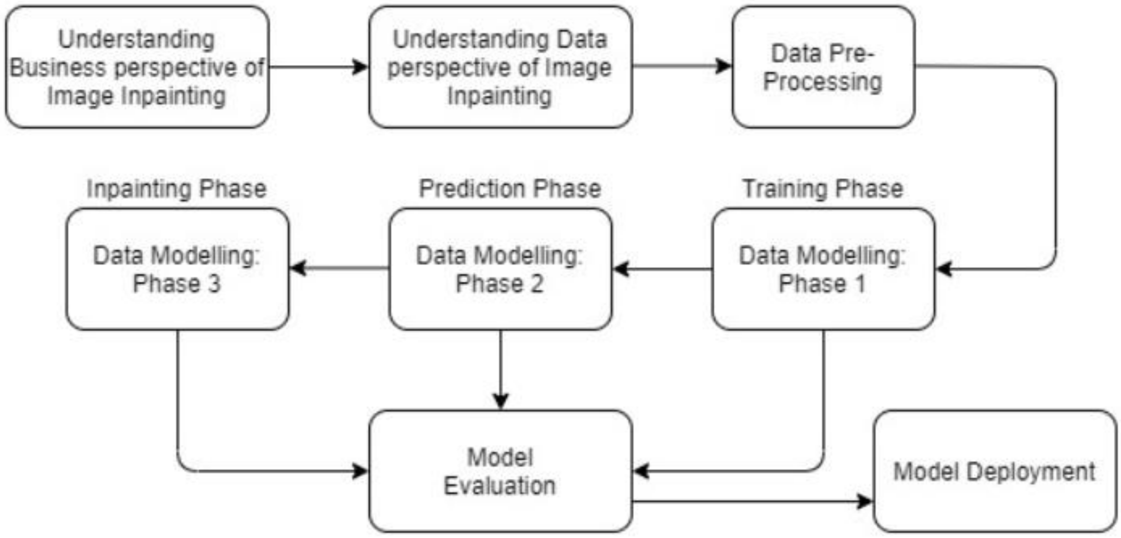


Figure 2: Overview of CRISP-DM Model[2]

3.1 Business Understanding

The topic of Image Inpainting is selected by taking its business point-of-view into account. The research has set objectives accordingly in order to tackle the challenges which are present in current techniques and to meet business goals as well as technical goals for the project. Business goals revolves around the truthfulness of Image Generation and its relationship with the domain related to the data used with achieving Business success criteria. Technical goals cover the desired requirements and the constraints for the research in achieving qualitative results with Data Mining success criteria. This research has generated unique representation of images during the training phase and reconstructed images from corrupt images in the inpainting phase by taking both goals into consideration.

3.2 Data Understanding

In total three datasets have been used in this research. The first dataset consists of 202,599 celebrity face images with all unique images. The second dataset consists 8,189 unique flower images and the final dataset consists of available 809 Pokémon images. All datasets are publicly available and are cleared of ethical concerns, if any. These datasets have been selected based on their relation within their images. Images from the celebrity faces dataset are closely related to each other due to the fact that the contextual and perceptual information of faces are nearly comparable. Images for Flowers dataset does have close relations in case of the Contextual information whereas the images from the Pokémon datasets are sparsely related due to the presence of unique structures within every image. Thus, the research focuses on evaluation of all three datasets and study about the nature of images in context of both contextual as well as perceptual information.

- **Contextual Information** refers to how the surrounding pixels of the image might be able to provide information about the missing pixel.
- **Perceptual Information** refers about the information that how close is the generated image to the real image in case of resemblance.

[2] Research flowchart is self-drawn after understanding the process.

A good value of both will provide good results for the reconstructed image. Also, upon evaluation of datasets, the research has also been able to implement Image Prediction. After carrying out the Data Quality Report for all three datasets, it is noted that the Celeb-A Faces Dataset and the Flowers Dataset have all unique values but also have irrelevant information in their edge pixels (E.g. Hat, Clothes, Leaf, etc.). Thus, Image Resizing is required to extract relevant information.
Thus, Business Understanding and Data Understanding formed a base to perform pre-processing and implementation which are explained in the next section.

4 Design Specification

Design Specification basically focuses on the flow in which this research is conducted. The flow is as follows.

- Datasets are gathered and three required datasets are selected based on their characteristics and project needs.
- For making use of the GPU, a Cuda Environment is created and the machine is loaded with Python.
- For using the Python, this research opted for Anaconda based-Jupyter Notebooks and all datasets are loaded in Python.
- For performing Neural Networks, this research focuses on implementing the PyTorch module.
- Next step involves installing required libraries and carrying out pre-processing which deals with cleaning, resizing and splitting the data for train and test samples.
- Following to the pre-processing stage, model implementation is carried out using the DCGAN and is executed in three phases with storing its output and losses for all three datasets.
- Final step includes performing Evaluation of the DCGAN model for its predicted images using the Inception Score metrics.

A brief description of pre-processing and model implementation is provided in the next section.

5 Implementation

5.1 Data Pre-Processing

Qualitative data pre-processing is the most important stage while undertaking a research. Data pre-processing is the procedure of converting the raw input data into a cleaned data for performing data modelling operations. It involves cleaning of data by handling missing or null values, corrupt data, removing irrelevant data by performing feature extraction or feature selection, and creating new variables required from the existing data. Below procedures are carried out in this research in order to obtain cleaned data for modelling purpose.

- Image Resizing

As mentioned in the previous section, images of Celeb-A faces dataset and the Flowers datasets are cropped and resized to remove irrelevant edge information from original images. Images are center cropped for 108x108 pixels in order to focus only on the information which is required.

- Image Conversion

To follow same image types, all Images files are converted to .jpg extension. This is carried out because the images other than .jpg extension have an additional alpha channel per pixel, thus making a total of four channels instead of three.

- Image Splitting

Prior to performing Data Modelling, the data is required to be split into training and validation samples. An 80:20 split is performed on all datasets as 80% of images are used for training the model and 20% are used to test and validate the model along with performing Image Inpainting.

Upon pre-processing, it is noted that there are no corrupt images in any of the datasets. As there are no classes within the datasets, exploratory analysis is not carried out. Instead, all the datasets are provided to the tensor and normalization is performed before providing them to the Data-Loader for carrying out modelling which will be explained in-detail in the next section.

5.2 Data Modelling

Data Modelling is an integral phase of this research methodology where the Data Mining models are implemented and executed. The modelling structure of this research is divided into three sections.

5.2.1 Deep Convolutional Generative Adversarial Networks (DCGAN)

Deep Convolutional GAN or DCGAN is an adaptation to the original Generative Adversarial Networks (GAN) model and one of the most successful network implementation for the Generative Networks (Radford, Metz and Chintala, 2016). It focuses on adversarial training between two Neural Networks, Discriminator and Generator.

1. Discriminator Network

Discriminator Network used in this research is a Convolutional Neural Network (CNN) as a Classifier. All the max pooling layers are replaced by strided convolutional layers. This network acts as a down-sampling network with real images as its one input and the out of the Generator Network as its other, providing a binary classifier value at its output. It does not contain any fully connected layer and used the LeakyReLU activation function in all the layers except for the output layer which uses a sigmoid function. Finally, this network uses Batch Normalization excluding the input layer of the network.

In this research, the Discriminator Network works as binary classifier to compare the output of the Generator network with the input images. The network is provided with input images and is trained on them to evaluate the output of the Generator network. It receives another input from the Generator and compares the two inputs to decide which input is real. It then backpropagates the output to the Generator in order to improve its training. The network has its own loss function which it tries to decrease.

2. Generator Network

Generator Network in this research is also a Convolutional Neural Network (CNN) with changes inside the layers. All the max pooling layers here are also replaced by strided convolutional layers. This network makes use of transposed convolution for performing up-sampling and the fully connected layers from CNN have been totally removed. The activation function used in this network for all layers is ReLU function apart from output which makes use of tanh function. Finally, this network uses Batch Normalization excluding the output layer.

In this research, Generator Network is used to create new unique images for all three datasets. A noise vector with a value of 100 is provided as input to the image and it created a noise image at the output. As, the network is connected to the Discriminator Network, the output the Discriminator Network is backpropagated to the Generator which helps to train the network for generating images like the original Input images with increased number of iterations. The network is incorporated with a loss function which tries to reduce the loss against the Discriminator.

Weights are initialized in both networks in order to transform input images within the hidden layers. These networks are trained using various hyperparameters which are tuned to optimize both networks. They are trained with mini-batch stochastic gradient descent (SGD) in order to avoid the situation of Mode-Collapse with a mini-batch size of 128. For optimizing the tuning of hyperparameters, the research has made use of Adam optimizers as they are widely used and most reliable optimizers. Learning Rate and Betas ($\beta 1$, $\beta 2$) are the parameters for optimizer which are set to 0.0002 and (0.5,0.999) respectively after adjusting their values. In order to track and calculate the losses for both networks, Binary Cross-Entropy Loss function is used. During the training period states of both the neural networks are saved for using them as pre-trained networks in the Prediction and the Inpainting models. In support to modelling both networks, original images are resized to 64x64 pixels and the feature values are updated to 128. Feature values are those which are passed to the convolutional layers along with the input image for training. Original paper had set them to 64. Finally, the number of iterations required for training is initialized. As all datasets have different number of images and different features, different iteration values have ben initialized for all datasets and the values are inversely proportional to the number of images in datasets.

5.2.2 Image Prediction Model

Image Prediction model is the second step in the process of Data Modelling for this research. It is a novel implementation and works as a bridge between training and inpainting methods. This research aims to perform Image prediction on all datasets based on the training model to generate predicted images. Then the generated images are used as parameters to evaluate the prediction model.
This research uses the loaded state of the Generator network as a part of the Image Prediction model. It also takes the Adam Optimizer state into consideration while performing Image Prediction. The model is evaluated with the use of the Inception Score (IS) metric. A high value of Inception Score determines the quality of the evaluation on the model. Inception Score is calculated for all three datasets and are compared with each other. Explanation of the comparison is provided in the Evaluation section below. Thus, based on Image Prediction and the quality of the Image, it becomes possible to understand whether the training model is overfitting, underfitting, working well or has been trapped in the Mode-Collapse state.

5.2.3 Image Inpainting Model

The final step of Data Modelling in this research is the implementation of the Inpainting model. Image Inpainting is an application of Computer Vision and the Image Inpainting model is used as an application extension to the DCGAN model in order to reconstruct images from a set of corrupt images. The model first prepares a mask on the provided input images, reshapes it as per requirement and then the mask is concealed on the input image. Just like the training procedure, a model is iterated in order to regenerate the masked region based on the image’s Contextual and Perceptual Information. Thus, these two terms introduce two new losses for the Inpainting model, and the model aims to reduce both losses to achieve realistic inpainting results.
Just like the previous model, this model also uses the loaded state of the Generator Network, but it also uses the loaded state of the Discriminator Network to act as a classifier to deciding whether the reconstructed images are real or fake. A new hyperparameter is introduced in this model known as Lambda (λ) which is coefficient for calculating the contextual and the

perceptual losses. It has been set to 0.1 and the Binary Cross-Entropy Loss has also been taken into consideration for tracking the overall model loss. Input for this model is from the validation samples from each dataset. Masks are created by multiplying the mask value with the input image and the mask position is scaled to the center of the image with a size of 0.3. Just like previous models, the inpainting model is also iterated for restoring images from corrupt masked images. This research carried out 2000 iterations for achieving refined results. Discussions about the reconstructed results are clarified in the next section.

Model Implementations and Evaluations have been performed on a PC system with specifications, Windows 10 Home environment, Intel® Core™ i5-9300H CPU @ 2.40GHz Processor, 8.00 GB RAM, NVIDIA GeForce GTX 1650 4GB Graphics. Code has been implemented using Python Programming Language in Anaconda Environment-built Jupyter Notebook and integrated with CUDA using NVIDIA cuDNN environment and Pytorch. Entire coding is performed on the GPU. Three experiments on three datasets respectively are conducted to achieve desired results which are described in-detail in the Evaluation section.

6 Evaluation

Evaluation is the most critical phase in the methodology as this phase provides a ground support to the work which is performed under modelling and helps to assess different test cases to reach a conclusion. This research focuses to compare three different scenarios in images from training, prediction and inpainting and to discuss the results generated from evaluations. Three scenarios are based the complexity of images are discussed in below experiments.

6.1 Experiment 1

Implementing DCGAN, Prediction and Inpainting on Celeb-A Faces Dataset.

This experiment is based on 1st scenario where Celeb-A Faces Dataset[3] is selected. The images in this dataset have a similar pattern (face images) and the features are also same (e.g. eyes, nose, mouth, etc.) with a small difference considering shades and extra features (e.g. hat, eyeglasses, etc.) The aim of using this dataset is to achieve better results than the other two because of its image nature.

In the pre-processing phase, a split is performed on the dataset to obtain training and validation samples and the samples are center-cropped and resized to 108x108 size to remove the redundant information as mentioned in the section 3.2. After performing data preparation, the cleaned data is passed to the Data Loader class to create minibatches for iterations. During training of the DCGAN model, a flow is followed.

1. Weights of Discriminator are initialized and updated to zero for avoiding over-fitting.
2. Discriminator is trained on real images from the dataset.
3. Discriminator is trained on fake images provided by the Generator.
4. Backpropagating the total error.
5. Updating weights of the Generator.
6. Training Generator on Noise vector and backpropagated input from Discriminator.

This process goes on iterating for 20 epochs and the model is trained to generate similar face images as the input images. Also, the neural network losses are tracked and are plotted to examine the training. Figure 2 exhibits the graph of network losses.

[3] http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html [Accessed on 10th August 2020]

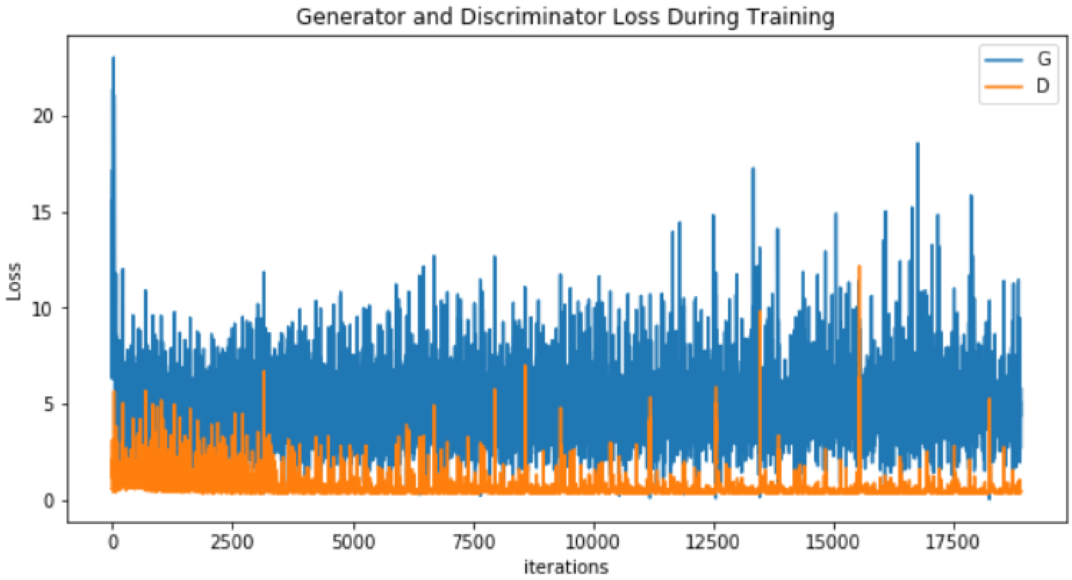


Figure 3: Generator and Discriminator Loss comparison for Celeb-A Faces Dataset

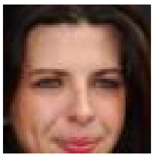
From Figure 2, the loss values can be understood. As noise vector is provided as input to the Generator, it obviously has a very high loss at the start, eventually leading to reduce the loss by training from backpropagated input from the Discriminator. There are few spikes in the Discriminator loss due to the ability of GAN to be very unstable. Stability of GAN has not been achieved yet globally, which is aimed to achieve in the future. The average Discriminator loss achieved is 0.4072 and the average Generator loss is 4.3868 which are good values as the Generator network keeps on improving itself to match and overtake Discriminator network.

Moving forward, the state of Generator is passed to the Prediction model as mentioned in section 4.2 to perform Image Prediction based on trained images. Upon execution, the predicted image is generated and is presented in Figure 3.

Figure 4: Predicted Images for Celeb-A Faces dataset (Rotated 90° Clockwise)

To evaluate the Predicted Image with respect to trained images, Inception Score (IS) is used as an Evaluation Metric. The reason of using Inception Score over other techniques like Geometry Score or Fréchet Inception Distance (FID) score, is because IS is robust to the critical Mode-Collapse situation which lacked by other techniques. Instead of going for pre-developed Inception v3 model to calculate the IS, this research has used KL Divergence formula. Upon evaluation the Image received the IS value of 33685.228103. It is a very good value which directly points to high quality of Image Prediction.

Final step refers to performing Image Inpainting. Validation samples are provided to perform inpainting based on states reproduced from the training phase. Input image, Masked Image and Inpainted images for the input are displayed below.



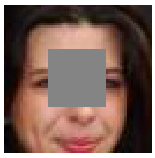


Figure 5: Source and Masked Face Images

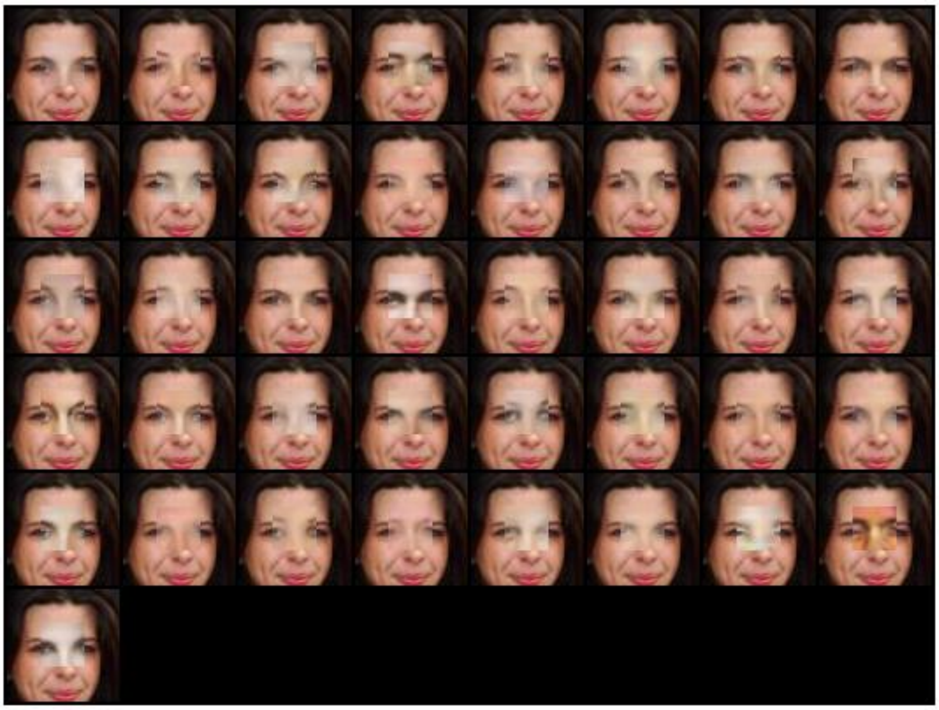


Figure 6: Inpainted Face Images

It can be observed from above images that the model is able to regenerate the missing pixels in most of the scenarios reducing both the context loss and the perceptual loss. The comparison of losses from 1st iteration with the final iteration is provided below.

Table 1: Loss Comparison for Celeb-A Faces Dataset

Iteration	Context Loss	Perceptual Loss
1	176137.3750	6.865256
2000	101073.0312	6.235963

From this experiment it is observed that good training, prediction and inpainting results are obtained for the Celeb-A faces dataset. This is due to the similarity in nature of the dataset as mentioned earlier. There are some uncertainties related to few generated images, but the images generated are better with the implementation of the Prediction model. The next experiment will focus on the 2nd scenario of the complexity of image.

6.2 Experiment 2

Implementing DCGAN, Prediction and Inpainting on Flowers Dataset.

This experiment performed on the Flowers Dataset[4] which has unique images of flowers. The 2nd scenario is considered as images of the dataset also have similar patterns, but attributes of images are different like shade and shape of flowers. So, these conditions increase the complexity for training, prediction and inpainting. The aim of this experiment is to achieve similar results as Experiment 1 despite of increase in complexity.

The pre-processing stage is similar to Experiment 1 and cleaned images are obtained after performing transformation. The modelling stage also follows similar procedure resulting to training the model for 100 iterations with flower images tracking network losses and plotting using a graph which is displayed below.

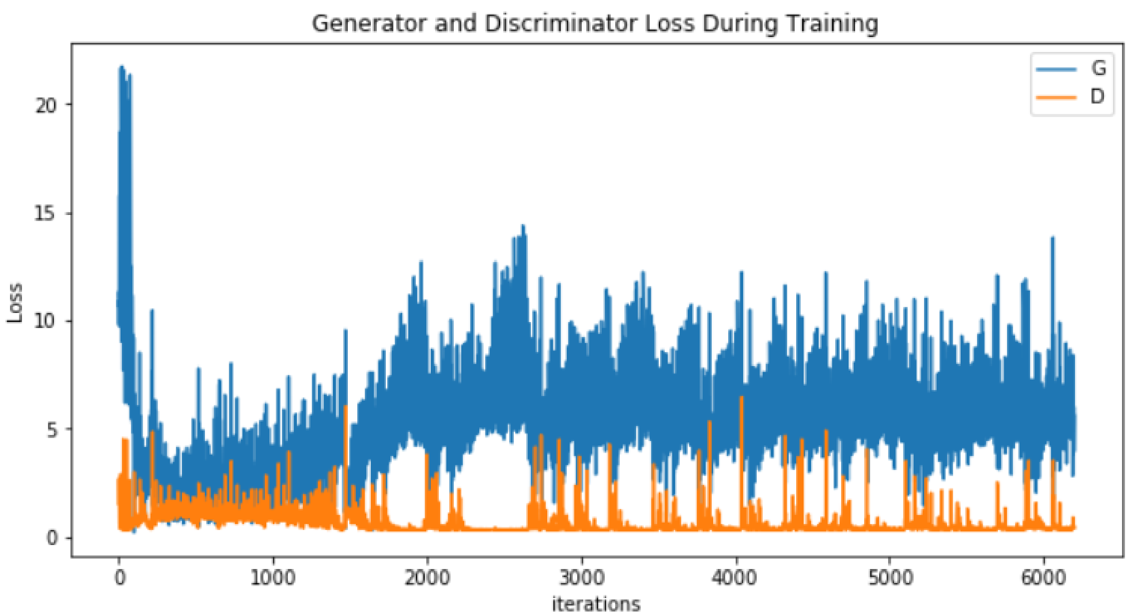


Figure 7: Generator and Discriminator Loss comparison for Flowers Dataset

Figure 6 displays the loss graph. It can be stated that due to the introduction of complexity, generated images are less similar to the input images but do exhibit nature of images. The Generator loss have more spikes, but it keeps on training the network to improve loss values.

[4] http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html [Accessed on 10th August 2020]

The obtained average Discriminator loss after training is 0.3988 and the average Generator loss is 5.9650.

Figure 8: Predicted Images for Flower Dataset (Rotated 90° Clockwise)

The Prediction model also follows the same procedure resulting into generating a predicted image based on loaded sate of the Generator network which is displayed in Figure 6. Image Prediction is a bit off-track in relation to this scenario due to intensive shade variations within images. Upon calculating the Inception Score of the predicted image, the value achieved is 25027.5135574 which is pretty good.
The final evaluation is performed on the Inpainting model following similar steps and inpainted images are presented below.

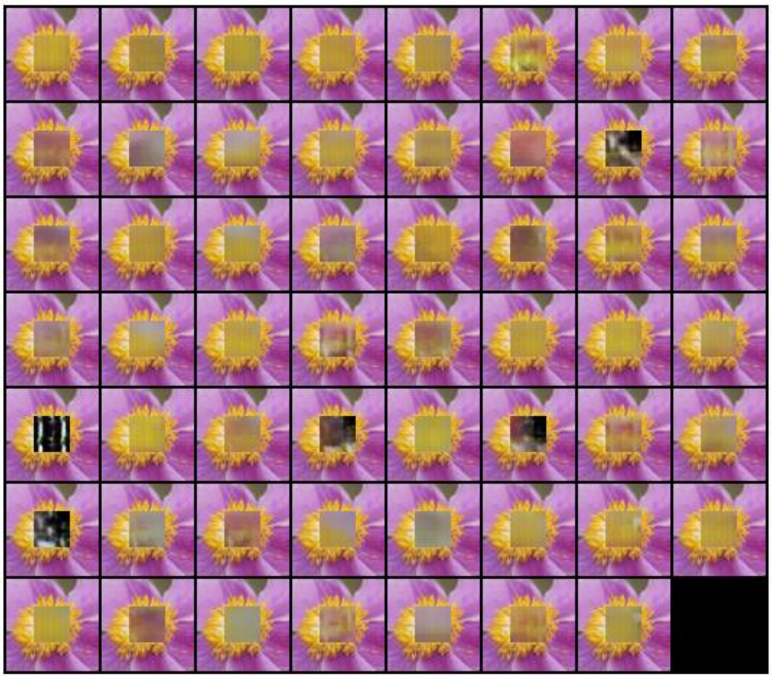


Figure 9: Inpainted Flower Images

Above images states that the Inpainting model can reconstruct images but not entirely as the images lack context information for achieving well restored image. Below table exhibits the context and the perceptual losses for 1st and the last iteration of Inpainting.

Table 2: Loss Comparison for Flowers Dataset

Iteration	Context Loss	Perceptual Loss
1	272758.0312	2.468180
2000	173338.6875	1.481369

This experiment demonstrates that Training and Inpainting Results are moderately good, but Prediction results are a bit above average for the Flowers dataset. This is due to the complexity of attributes within images of the dataset as mentioned earlier. There are bit more uncertainties related to few generated images, but the images generated are better with the implementation of the Prediction model. The next experiment will focus on the final scenario of the complexity of image.

6.3 Experiment 3

Implementing DCGAN, Prediction and Inpainting on Pokémon Dataset.

The final experiment is performed on the dataset which includes images of 809 unique Pokémon[5]. Yes, the Pokémon universe has really expanded a lot! Jokes aside, the decision of selecting the dataset is hugely because of the diversity of Images. All images are inimitable with different features, including shape, size, color, etc. leading to further increase the complexity of this experiment. So, there no relation to contextual and perceptual information within images.

The training phase is executed like other experiments with similar functions and the images are trained for around 500 epochs due to a smaller number of images in the dataset. The graph of loss functions related to training are displayed below in Figure 8.

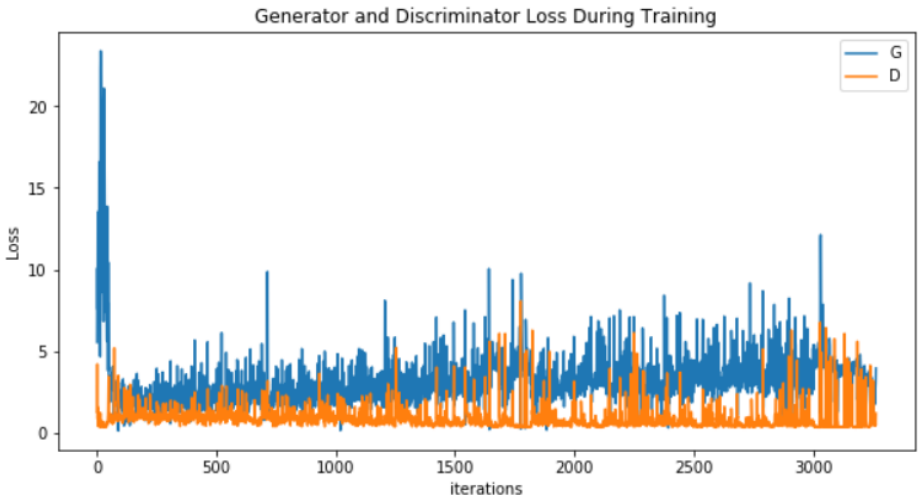


Figure 10: Generator and Discriminator Loss comparison for Pokémon Dataset

[5] https://www.kaggle.com/kvpratama/pokemon-images-dataset [Accessed on 10th August 2020]

From the graph, it is observed that the training performed on neural networks are very closely matched resulting into low loss values. There are spikes in Generator network training due to unstable nature of GAN as stated earlier but the images do provide good visualizations of generated images from training and might be considered for creating more Pokémon in the future! The obtained average Discriminator Loss value is 0.4798 and the average Generator Loss value is 3.1275.

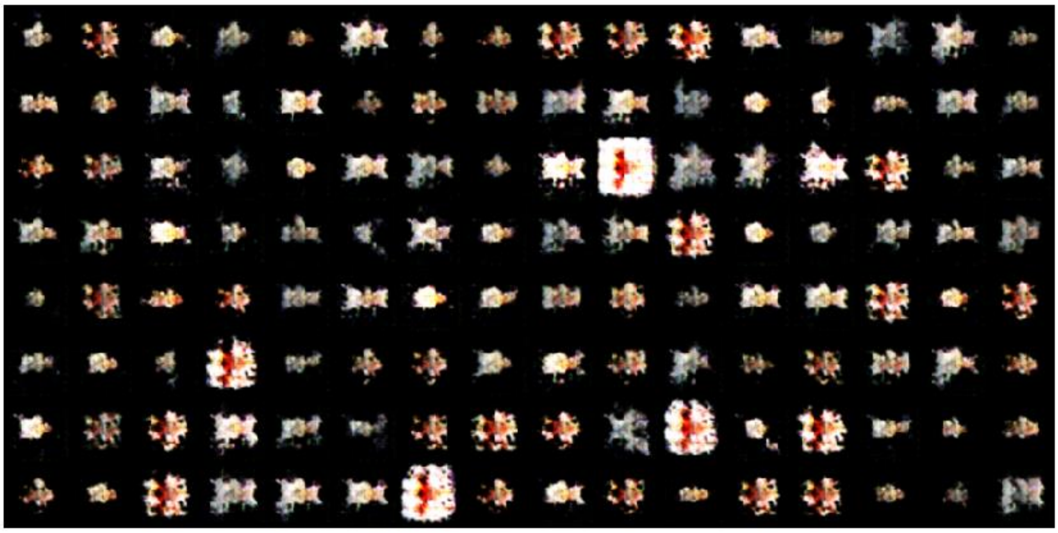


Figure 11: Predicted Images for Pokémon Dataset (Rotated 90° Clockwise)

The parameters of the Prediction model are like previous experiments and the model can generate pretty good images from prediction which are displayed in Figure 9. Evaluation of the predicted image resulted in the Inception value of 109669.928566 which is indeed better than the other two experiments.

The main challenge for this experiment is the Inpainting task as the images lack both context as well as perceptual information. The model inherits similar parameters and upon performing the inpainting, below images were generated.

Figure 12: Inpainted Pokémon Images

Form the above figure, it can be clearly observed that due to the complexity of the nature of the Pokémon images, the model is not able to gather enough data to the perform Inpainting task.
Thus, this experiment validates that Training and Prediction Results are good, but the model fails to generate good Inpainting results due to dataset's natural complexity. Uncertainties are present in each modelling phase, but the images generated are better with the implementation of the Prediction model. This research aims to work on this limitation thoroughly and implement a fully robust prediction and inpainting model in the future.

7 Conclusion and Future Work

The DCGAN model is used in a wide variety of applications in order to perform Image Processing. One such application is Image Inpainting, in which the model implementation has obtained a huge success over the years with certain modifications. The novel use of implementing Image Prediction for carrying out Inpainting has proven to be quite helpful in generating images in this research. The implementation of Image Prediction model really proved to a bridge between the Training model and the Inpainting model by evaluating images and deciding the amount of training needed to obtain realistic results.
Upon performing experiments on three datasets based on their nature of complexity, it is observed that images from the Celeb-A Faces dataset have been successfully predicted and inpainted with a good Inception Score than the other two datasets. This has been achieved due to the presence of both Contextual and Perceptual Information within images. Even though Flowers dataset and the Pokémon dataset lacked in producing Predicted images and Inpainted images respectively, all experiments were good enough in supporting Image Prediction.
This research aims to develop a robust Prediction and Inpainting model in future that will be able to reconstruct images for all three scenarios presented in the research. Also, the research aims to expand its scope to support and regenerate super-resolution images in the future. Since implemented models are able to perform good visual image generation, image prediction and image inpainting for the Celeb-A Faces dataset and considering the Global Pandemic situation into consideration, the research intends to develop greater masked images and a robust Face inpainting model in security domain to combat theft.

References

Anirudh, R. *et al.* (2019) 'Improving Limited Angle CT Reconstruction with a Robust GAN Prior', *NeurIPS 2019 Workshop on Deep Inverse Problems*, pp. 1–5. doi: arXiv:1910.01634.

Cai, W. and Wei, Z. (2020) 'PiiGAN: Generative Adversarial Networks for Pluralistic Image Inpainting', *IEEE Access*. IEEE, 8, pp. 48451–48463. doi: 10.1109/ACCESS.2020.2979348.

Demir, U. and Unal, G. (2018) 'Patch-Based Image Inpainting with Generative Adversarial Networks'. doi: arXiv:1803.07422.

Goodfellow, I. J. *et al.* (2014) 'Generative adversarial nets', *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, pp. 2672–2680. doi: 10.5555/2969033.2969125.

Hirvola, O. *et al.* (2016) 'Improved image quality in fast inpainting with omnidirectional filling', *2016 International Conference on Image, Vision and Computing, ICIVC 2016*. IEEE, pp. 31–35. doi: 10.1109/ICIVC.2016.7571269.

Huang, X. *et al.* (2017) ‘Stacked generative adversarial networks’, *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January, pp. 1866–1875. doi: 10.1109/CVPR.2017.202.

Hui, Z. *et al.* (2020) ‘Image Fine-grained Inpainting’, doi: arXiv:2002.02609

Ishi, M. S., Singh, L. and Agrawal, M. (2015) ‘Reconstruction of images with exemplar based image inpainting and patch propagation’, *2014 International Conference on Information Communication and Embedded Systems, ICICES 2014*. IEEE, (978), pp. 1–5. doi: 10.1109/ICICES.2014.7034051.

Karras, T., Laine, S. and Aila, T. (2019) ‘A style-based generator architecture for generative adversarial networks’, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June, pp. 4396–4405. doi: 10.1109/CVPR.2019.00453.

Li, A. *et al.* (2019) ‘Boosted GAN with Semantically Interpretable Information for Image Inpainting’, *Proceedings of the International Joint Conference on Neural Networks*. IEEE, 2019-July(July), pp. 1–8. doi: 10.1109/IJCNN.2019.8851926.

Liu, G. *et al.* (2018) ‘Image Inpainting for Irregular Holes Using Partial Convolutions’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11215 LNCS, pp. 89–105. doi: 10.1007/978-3-030-01252-6_6.

Liu, H. *et al.* (2018) ‘Image inpainting based on generative adversarial networks’, *ICNC-FSKD 2018 - 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*. IEEE, pp. 373–378. doi: 10.1109/FSKD.2018.8686914.

Liu, M. Y. and Tuzel, O. (2016) ‘Coupled generative adversarial networks’, *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pp. 469–477. doi: 10.5555/3157096.3157149.

Mao, X. *et al.* (2017) ‘Least Squares Generative Adversarial Networks’, *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October, pp. 2813–2821. doi: 10.1109/ICCV.2017.304.

Nazeri, K. *et al.* (2019) ‘EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning’. doi: arXiv:1901.00212.

Radford, A., Metz, L. and Chintala, S. (2016) ‘Unsupervised representation learning with deep convolutional generative adversarial networks’, *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pp. 1–16. doi: arXiv:1511.06434.

Salimans, T. *et al.* (2016) ‘Improved techniques for training GANs’, *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 2234–2242. doi: 10.5555/3157096.3157346.

Vitoria, P., Sintes, J. and Ballester, C. (2019) ‘Semantic image inpainting through improved Wasserstein generative adversarial networks’, *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 249–260. doi: 10.5220/0007367902490260.

Xu, S., Liu, D. and Xiong, Z. (2018) ‘Edge-guided generative adversarial network for image inpainting’, *2017 IEEE Visual Communications and Image Processing, VCIP 2017*, 2018-January, pp. 1–4. doi: 10.1109/VCIP.2017.8305138.

Xu, Z. and Sun, J. (2010) ‘Image inpainting by patch propagation using patch sparsity’, *IEEE Transactions on Image Processing*, pp. 1153–1165. doi: 10.1109/TIP.2010.2042098.

Yang, C. *et al.* (2017) ‘High-resolution image inpainting using multi-scale neural patch synthesis’, *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 4076–4084. doi: 10.1109/CVPR.2017.434.

Yang, D. *et al.* (2019) ‘Diversity-sensitive conditional generative adversarial networks’, *7th International Conference on Learning Representations, ICLR 2019*, pp. 1–23.doi: arXiv:1901.09024.

Yeh, R. A. *et al.* (2017) ‘Semantic image inpainting with deep generative models’, *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January, pp. 6882–6890. doi: 10.1109/CVPR.2017.728.

Yu, J. *et al.* (2018) ‘Generative Image Inpainting with Contextual Attention’, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5505–5514. doi: 10.1109/CVPR.2018.00577.

Zhang, H. *et al.* (2019) ‘Self-attention generative adversarial networks’, *36th International Conference on Machine Learning, ICML 2019*, 2019-June(May), pp. 12744–12753. doi: arXiv:1805.08318.