IMAGE OUTPAINTING USING DEEP GENERATIVE MODELS

A Project Report

Submitted by

Y. SAI AKHIL [CB.EN.U4CSE17050]
ILAM. PRATHYUSHA [CB.EN.U4CSE17424]
T. P. V. KRISHNA TEJA [CB.EN.U4CSE17465]
K. AJAY KUMAR REDDY [CB.EN.U4CSE17628]

Under the guidance of

Dr. K. Raghesh Krishnan
(Assistant Professor (Sr.), Department of Computer Science & Engineering)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

Amrita Nagar PO, Coimbatore - 641 112, Tamilnadu

May 2021

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF ENGINEERING, COIMBATORE – 641 112

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **IMAGE OUTPAINTING USING DEEP GENERATIVE MODELS** is submitted by Sai Akhil Y(cb.en.u4cse17050),Ilam Prathyusha(cb.en.u4cse17424),T.P.V Krishna Teja(cb.en.u4cse17465), K.Ajay Kumar Reddy(cb.en.u4cse17628) in partial fulllment of the requirements for the award of the Degree Bachelor of Technology in Computer Science and Engineering is a bonafide record of the work carried out under our guidance and supervision at Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore.

SIGNATURE

Dr. K. Raghesh Krishnan
PROJECT GUIDE
Assistant Professor (Sr.)
Dept. of Computer Science & Engineering

SIGNATURE

Dr. (Col.) P. N. Kumar
CHAIRPERSON
Dept. of Computer Science & Engineering

Signature of the Internal Examiner

Signature of the External Examiner

DECLARATION

We, the undersigned solemnly declare that the project report **IMAGE OUTPAINTING USING DEEP GENERATIVE MODELS** is based on our own work carried out during the course of our study under the supervision of Dr. K. Raghesh Krishnan, Assistant Professor (Sr.), Computer Science & Engineering, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgement have been made wherever the findings of others have been cited.

Y. SAI AKHIL[CB.EN.U4CSE17050]

ILAM. PRATHYUSHA [CB.EN.U4CSE17424]

T. P. V. KRISHNA TEJA[CB.EN.U4CSE17465]

K. AJAY KUMAR REDDY[CB.EN.U4CSE17628]

ABSTRACT

This work explores neural network models for extrapolating images or outpainting images given a static image of certain dimensions as input. Compared to image prediction models that are in existence, the image outpainting models are relatively less and also not a hundred percent accurate in predicting the outward image. Image outpainting has less background to catch in the image centre and more material to anticipate at the image boundary. As a result, current classical encoder-decoder models may not be able to accurately predict the outstretched unknown material. Hence, new architecture is proposed for the problem of image outpainting utilizing the power of deep generative modelling. This work also looks at the potential of the proposed model for a number of exciting applications that can aid researchers in different fields.

ACKNOWLEDGEMENTS

We would like to express our deep gratitude to our beloved Satguru **Sri Mata Amrita--nandamayi Devi** for providing the bright academic climate at this university, which has made this entire task appreciable. This acknowledgement is intended to be a thanks-giving measure to all those people involved directly or indirectly with our project. We would like to thank our Pro Chancellor **Bramachari Abhayamrita Chaitanya**, Vice Chancellor **Dr. Venkat Rangan. P** and **Dr. Sasangan Ramanathan**, Dean Engineering of Amrita Vishwa Vidyapeetham for providing us the necessary infrastructure required for the completion of the project. We express our thanks to **Dr. (Col) P.N. Kumar**, Chairperson of Department of Computer Science Engineering, **Dr. G. Jeyakumar and Dr. C. Shunmuga Velayutham,** Vice Chairpersons of the Department of Computer Science and Engineering for their valuable help and support during our study. We express our gratitude to our guide, **Dr. K. Raghesh Krishnan, Assistant Professor (Sr.)** for his guidance, support and supervision. We feel extremely grateful to **Dr. N. Radhika,Professor, Abirami K.,Assistant Professor,Ms. Neethu MR,Faculty Associate** for their feedback and encouragement which helped us to complete the project. We would also like to thank the entire staff of the Department of Computer Science and Engineering. We would like to extend my sincere thanks to our family and friends for helping and motivating me during the course of the project. Finally, we would like to thank all those who have helped, guided and encouraged me directly or indirectly during the project work. Last but not the least, we thank God for His blessings which made my project a success.

Names & Roll Nos of all team members:

Y. SAI AKHIL[CB.EN.U4CSE17050]

ILAM. PRATHYUSHA [CB.EN.U4CSE17424]

T. P. V. KRISHNA TEJA[CB.EN.U4CSE17465]

K. AJAY KUMAR REDDY[CB.EN.U4CSE17628]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

GAN	Generative Adverserial Network
SRN	Semantic Regeneration Network
RMSE	Root Mean Squared Error
MRF	Markov Random Field
DCGAN	Deep Convolutional Generative Adverserial Network
RCT	Recurrent Content Transfer
SHC	Skip Horizontal Connection
VAE	Variational Auto Encoder
MSE	Mean Squared Error

List of Symbols

α, β	Damping constants
θ	Angle of twist, rad
ω	Angular velocity, rad/s
b	Width of the beam, m
h	Height of the beam, m
$\{f(t)\}$	force vector
$[K^e]$	Element stiffness matrix
$[M^e]$	Element mass matrix
$\{q(t)\}$	Displacement vector
$\{\dot{q}(t)\}$	Velocity vector
$\{\ddot{q}(t)\}$	Acceleration vector

Chapter 1

INTRODUCTION

1.1 Background

With the advent of generative adversarial models, many developments have been made in the field of computer vision. One of them includes inpainting which aims to recover the damaged parts or quality of the image. In this case, inpainting requires a model which understands the context of the image and improves or recovers the image.This project also has a tedious task of outpainting which aims to recover the background for a foreground image.

1.2 Problem Definition

Given an input image, humans can easily see and visualize how the nearby elements in the image would have looked, if they had been captured. For example, given a picture of the mountains, human brain can try to picture an area surrounded by forests or snow, think of a lake below a hill, and it can visualize the cliffs near the sea. This mental capacity depends on their prior knowledge and exposure to various places. In other words, this is an image outpainting task. It can allow a variety of applications to create content such as editing an image using advanced regions, panorama image production, and extended experience, to name a few.

1.2.1 Motivation

The recent developments in image inpainting do not directly address the outpainting problem as the former has a lot of context to deal with the missing pixels have a larger amount of surrounding pixels, acting as boundary conditions and provide an important guide to inpainting. On the contrary, the problem of outpainting can only depend on the context of the available image, there are only a few number of pixels near the boundary available as the boundary condition. In addition, the texture and semantics of the drawn

regions must match those of input. Finally, drawing methods should support diversity of the produced content. The same similarity is between video interpolation and video prediction, where the former works with existing events while the latter attempts to model variable futures.

Chapter 2

LITERATURE SURVEY

The following section provides a review of the literature related to image extrapolation and outpainting:

Wang et al. (2019) aim to solve the problem of image extrapolation[5] by utilizing the power of deep generative models.The authors have proposed a Semantic Regeneration network[5] to which they have added several special modules and have used a variety of spatial related losses. The Semantic Regeneration Network (SRN) [5] consists of two networks where the first network extracts complex inferences from the given image and a second network that upsamples these features into images. The authors have proposed a context normalization(CN)[5] module to enhance the quality of the generated image. The authors have used Relative Spatial Variant Loss[5], Implicit Diversified Markov Random Field (MRF) Loss[5] and Context Adverserial Loss[5] as the three losses and the combined loss to minimize. The network was tested on datasets lke CelebA-HQ, CUB200, DeepFashion[5], Paris street view[5]. In the end,the authors were able to produce consistent images with high-quality textures.

Sabini and Rusak (2018) aim to improve the existing solutions to image outpainting (also called as image extrapolation) by proposing a three-phase training process which involves training a Deep Convolutional Generative Adverserial Network (DCGAN)[4] architecture on a part of the Places365[4] data set. The authors have also proposed a way to enhance the quality of the outpainted image by using a local discriminator and increased dilated convolutions. The first phase in the process is training the generator (by updating its weights). The second phase trains the discriminator and in the third phase, the generator and discriminator are adversarially trained. The authors have considered Root Mean Squared Error (RMSE)[4] as the quantitative metric. The use of local discriminators reduced the RMSE and improved color fidelity[4], at the cost of increased training time. Also, the authors have experimented with different dilation

rates[4] for the dilated convolution layers of the generator and found that the network was able to reconstruct the original image properly with increased dilation. The authors have been able to produce 128*128 realistic color images. Along with this, an investiga--tion into recursive[4] outpainting was proposed which is indiscriminately extending an image beyond the ground truth image. Although there was a lot of noise stacking up as the iterations increase, the outpainted image was comparatively realistic. The output of recursive outpainting model is depicted in Figure 2.1.

Figure 2.1: Each right image is the result of recursively outpainting the corresponding left image five times.

In this paper Xiaofeng Zhang (2020), to introduce a task-specific architecture, the authors proposed a two-stage siamese adversarial[6] extrapolation model for images which is known as Siamese Expansion Network (SiENet)[6](refer Figure 2.2). SiENet[6] is designed by introducing boundary-sensitive convolution called adaptive filling convo--lution[6] and is designed to automatically detect the features of surrounding pixels that are outside the known content along with balance in smoothness and characteristic. This adaptive filling convolution[6] is provided into the encoder so that encoder could get information about the unknown content and this technique reduced the burden on decoder. This could help to predict pixels of unknown areas. The author used two sets of encoder-decoders, where the first set is used as a structure generator[6] and the second set is used as content generator[6]. The encoder of the structure generator

down samples the input with 8 scales. Two residual blocks[6] are used to capture multi-scale information and are designed to continue the encoder's output. Finally, the output of residual blocks is upsampled to the desired resolution using three nearest-neighbor interpolation. The content generator[6] has a similar comprehensive structure to the structure generator, with the exception of inserting two residual blocks before the final two upsampling operations. The metrics used for assessment are structural similarity (SSIM)[6] and peak signal-to-noise ratio (PSNR)[6]. The author uses the datasets cityscapes[6], paris street-view[6], beach and Scenery[7], which include street and nature instances.

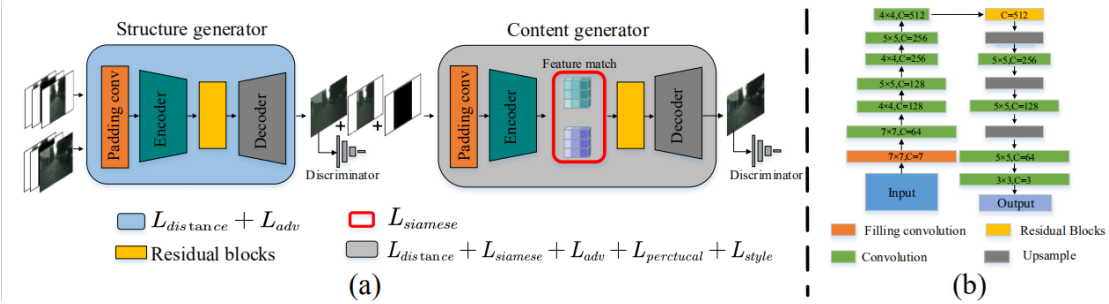

Figure 2.2: SiENet Architecture Diagram

The main emphasis of this paper Zongxin Yang (2019) is on producing very long images(refer Figure 2.3) that are spatially and semantically consistent[8] with the original input. Through fusing the encoder's information into the decoder, the encoder's infor--mation can be fully used. The author devised a Skip Horizontal Connection (SHC)[3] to bind encoder and decoder at the same level for this purpose. In this way, the decoder will make a strong prediction based on the data. In addition, the author suggested Recurrent Content Transfer (RCT)[3], which involves transferring the sequence from the encoder to the decoder in order to produce new contents. The author declared that Recurrent Content Transfer (RCT)[3] would facilitate the network compared to channel-wise maximum connection strategy. The author claimed that RCT[3] would help the network manage the spatial relationship in the horizontal direction more effec--tively than channel-wise full communication strategy. The output of the previous step is used as the input for the next step. As a result, it's known as a recursive operation, and it can theoretically generate smooth and realistic large-scale images. RCT outputs feature maps with the same dimension as the input feature maps, which are 4 x 4 x

1024[3]. Decoder generates a 128 x 256 image by combining 4 x 4 x 1024 dimensional features encoded from a 128 x 128 image. The architecture proposed predicts the right half of the created image, while the left half is identical to the input image. To increase the spatial size and reduce the channel number, the author proposed five transposed-convolutional layers[3] in the decoder. The author proposed using the built Skip Horizontal Connection (SHC)[3] to fuse the encoder function into the decoder in each transposed-convolutional layer. The only drawback of this model is that it can only accommodate fixed-size features. Masked reconstruction loss and an adversarial loss are the two aspects of the loss feature that are considered[3]. The author developed the dataset, which consists of 6000 diffraction-corrected images. The author have used the dataset, which consists of 6000 images of various and complex natural scenes, including mountains with and without snow, valleys[3], seaside[3], starry sky[3], riverbank[3], and so on.

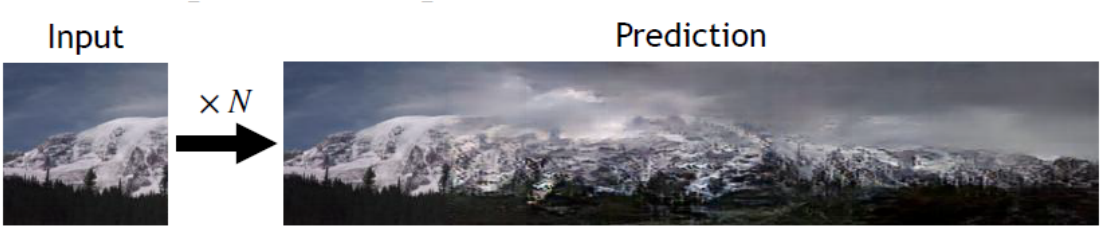

Figure 2.3: Results of very long image prediction

This paper Kyunghun Kim and Kang (2020) aims to outpaint using both outpainting and inpainting techniques. An edge map generation network and an image completion network are the two parts of this model. A generator-discriminator pair makes up each section. Ge and De are the generator and discriminator of the structural edge map generator, respectively. The global edge map image Epred, which is used in the image completion network, is named after the word Ge, which is used to predict missing structures. The picture completion network's generator and discriminator are Gc and Dc, respectively. By drawing information from the edge map image, Gc creates the final image, Ipred.

Zhang et al. (2019) aims to find the possible background for the given input fore--ground region. Generally, the tendency of humans to predict the possible background for a foreground. The same is trying to be achieved by this model. For example,

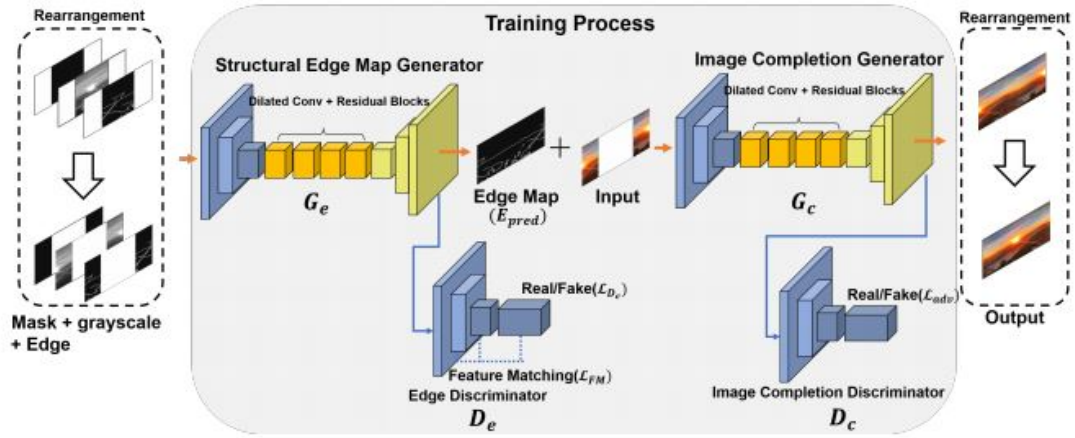

Figure 2.4: Architecture diagram

given the eyes or nose of a face as an input, the background can be different faces with different styles. Previous approaches include Variational Auto Encoder (VAE)[8] with Generative Adverserial Network (GAN)[8] where VAE[8] is used to encode a distribution of various possible images for the input and GAN[8] to synthesize the realistic image. The process used involves Normalized diversification, Diversity Regular--ization, and lastly, feature pyramid discriminator[8]. The architecture diagram of the model is depicted in Figure 2.4. This process aims to achieve greater accuracy comparatively. The results of this model is depicted in figure 2.5.

Figure 2.5: Results of Multimodal Image Outpainting

Xiao et al. (2020) proposed a novel image-to-image synthesis architecture[6] to deal with a challenging issue of image outpainting. Dense residual learning[6] was used in the design, which resulted in deep model efficiency and favorable predictions. Dense residual learning was used to build the extrapolation network. Encoder-decoder models are a popular starting point for image-to-image conversion models. This architecture was used to build the generative network. Unlike traditional encoders and decoders, the developed generative network is built with dense residual neural layers. In comparison

to conventional encoder-decoder models, the model has more depth. When there are consecutive downsamples in an encoder-decoder, important information is lost, which is not the case in this modern generative network. The semi-skip connecting method was used to overcome the flaws in traditional U-Net[6] architecture and pass both low and high-frequency information.The architecture of Semi U-Net[6] was proposed to solve the issue of excessive computation. Skip shortcuts are only introduced in Semi U-Net[6] between the early layers in the encoder where feature maps are down-sampled and the corresponding deconvolutional layers[6] in the decoder. The results of this model is depicted in figure 2.6.

Figure 2.6: Results obtained using image-to-image synthesis architecture

Gardias et al. (2020) focused onGAN architecture[1] which contains two major additions: the first is the use of residual blocks in the generator, and the second is the use of two discriminators. This GAN architecture[1] encodes the image's background before generating the resulting image using deconvolutional[1] layers. This is achieved while maintaining the consistency of the features. The generative network was subjected to a residual model, which involved first convolving the image into its feature space and then upsampling[1] it into its extended version using deconvolution[1]. In order to enhance image quality, the encoder portion of the generative network includes several residual blocks. Dual discriminators make up the discriminator network. The first discriminator takes the entire generated image as input and uses a convolutional[1] network to determine if the image is real or generated, while the second discriminator is local and only looks at the bounds between image and the hallucinations. The local discriminator's goal is to eliminate the obviously fake hallucinations[1] that arise from low-quality generations or manifestations, which are instantly obvious to the human eye[1]. The results obtained by using this model is depicted in figure 2.7.

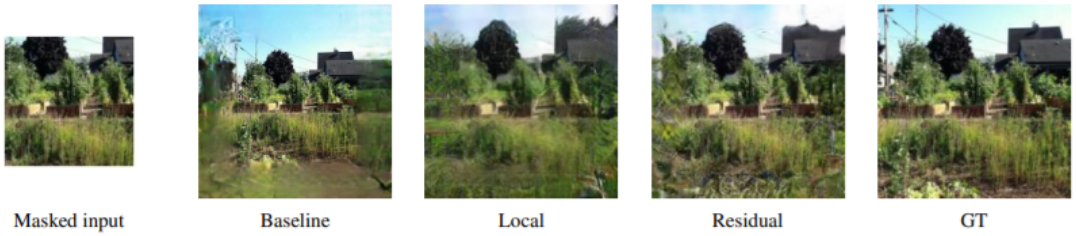

Figure 2.7: Results of Enhanced Residual Network

From all the above papers that are considered for literature survey, it has been observed that the model for image outpainting involves 2 modules: Generator and Discriminator, where the Generator is used to get the all possible background for a specific foreground and discriminator will select the appropriate background for the foreground.

2.1 Data Set

The proposed GAN model is evaluated on two types of datasets:

1. Landscapes dataset: This dataset contains images scraped from the Flickr website. The scraped dataset contains around 4300 images and splitted into train and validation datasets. This dataset is chosen as landscapes provide a great challenge for the model to outpaint the images as the model has to learn a lot of various different content in the image to properly outpaint it.

2. Cat2dog: 871 cat (birman) images, 1364 dog (husky, samoyed) images crawled and cropped from Google Images with dimensions 178*218. The cat2dog dataset has also been used in other tasks like Image-to-Image translation and hence has been proven to be a good benchmark dataset for both inpainting and outpainting tasks. The sample of the dataset is depicted in figure 2.8.

Figure 2.8: Sample images from the cat2dog dataset.

2.2 Software/Tools Requirements

1. Keras - A Python-based open-source neural-network library. It is user-friendly, scalable, and extensible, with the goal of allowing fast experimentation with deep neural networks.

2. Tensorflow - A free and open-source software library for dataflow and differentiable programming that can be used to solve a variety of problems. It's a symbolic math library that's also used in neural networks and other machine learning applications.

3. Pytorch - PyTorch is an open source machine learning library based on the Torch library, designed primarily by Facebook's AI Research lab for applications such as computer vision and natural language processing. It's open-source software distributed under the Modified BSD licence.

4. Google Colab - A Google Cloud-based Jupyter notebook environment that is completely free. Free GPU and TPU control, as well as Tensorflow/Pytorch integration.

Chapter 3

PROPOSED SYSTEM

3.1 System Analysis

3.1.1 System requirement analysis

Functional Requirements

- Our model should be able to efficiently outpaint any given image from our dataset.
- Any Python 3 installed computer with Pytorch version 1.8.1 or above.

Non-Functional Requirements

- To run the algorithm a user must need at least 4GB of GPU memory or any cloud computing or online service such as Google colaboratory.
- A decent and consistent internet connection is required if the model is being trained using an online service.

3.1.2 Module details

1. Preprocessing - Resize and applying mask
2. Generator
3. Discriminator
4. Postprocessing - Blending images

3.2 System Design

The proposed architecture leverages the power of GANs as they have proven to perform better at the task of image outpainting. The proposed GAN architecture comprises two innovative modules compared to general GAN: the first is the use of dense blocks in the generator, and secondly the use of multiple discriminators in the place of single discriminator to enhance the classifying ability of the discriminator and achieve better results.

3.2.1 Generator

The Generator aims first tries to understand the context and features in an image and then outputs an image having spacial and feature consistency using deconvolutional layers. First the image is passed as input to a dense block followed by a transition block which is followed by a series of convolutional layers to extract the features from the image and then reconstruct the image back to its original version by using deconvolutional layers(refer figure 3.1). Traditionally, dense blocks usually begin with the batch normalization layer but we have not used it in our model(refer Table 3.1) as we have found it is causing blurring of the outpainted image. The generator can be extended in the future to contain multiple dense blocks to enhance the image quality.

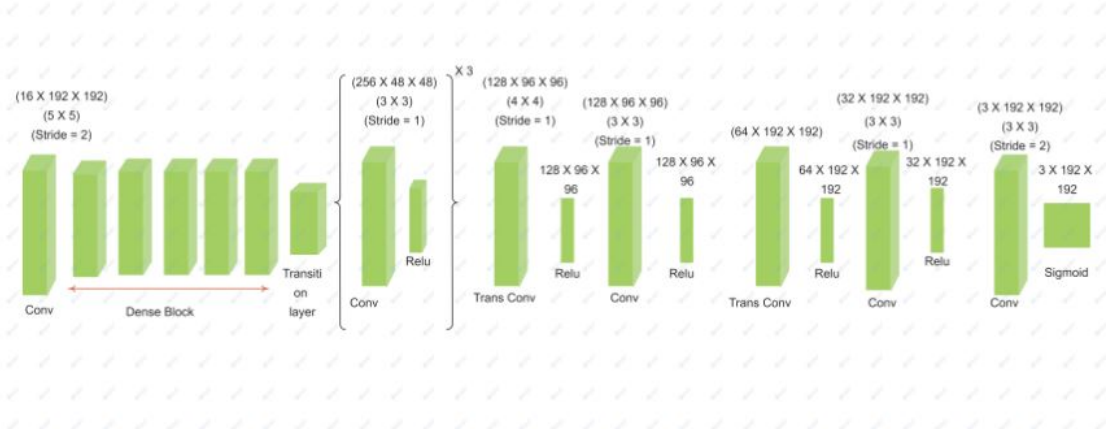

Figure 3.1: Generator Architecture

Layer	Parameters
BatchNorm2d	None
ReLU	None
Conv2d	k=3,stride=1
Conv2d	k=3,stride=1
Conv2d	k=3,stride=1
Conv2d	k=3,stride=1
Conv2d	k=3,stride=1

Table 3.1: Architecture of Dense Block

The dense block(in generator) applies the given input stride to the convolution layer. Here, the feature maps are concatenated from the previous layers and passed as input to the current layer.

3.2.2 Discriminator

The proposed model uses the concept of dual discriminator(refer figure 3.2). One of the discriminator is Global Discriminator, that takes the full outpainted image as input and passes it through a series of convolutional layers to classify whether the image is real or outpainted. The second is a Local Discriminator and is made to focus only on the patch generated between the input image and the outpainted part of the image. The outputs from both the discriminators are averaged. The main use of local discriminator is an attempt to eliminate the clearly distinguishable predicted images with ground truth, so to minimize the resulting images which are low in quality and can be easily identified as fake by the human eye.

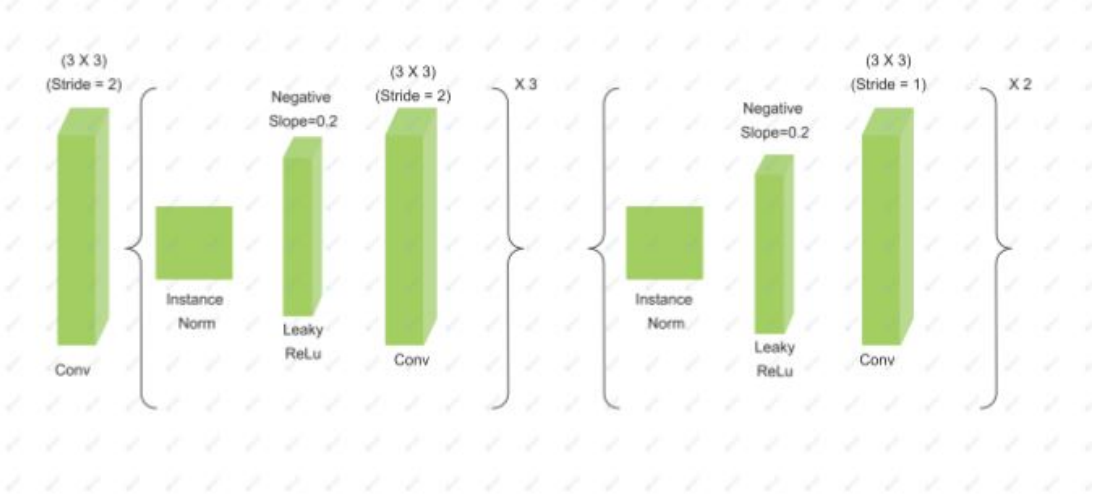

Figure 3.2: Local & Global Discriminator Architecture

3.2.3 Flow diagram of the system

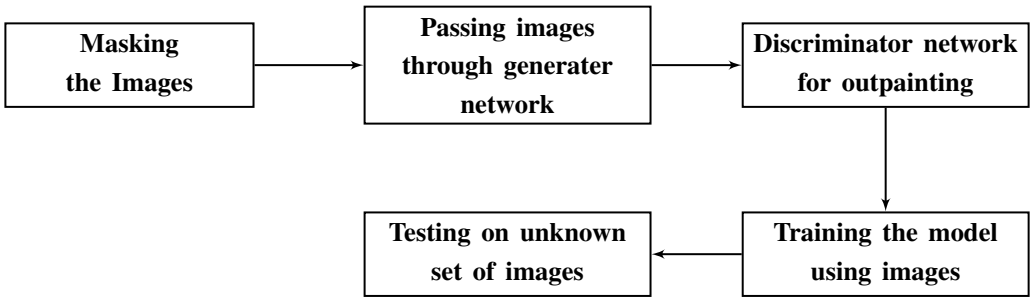

Figure 3.3: Overall approach

1. Images are first preprocessed to the specific size of 192x192.
2. Masking of the images is done so that only the center part of image of dimensions 128x128 is given as input to the Generator.
3. Generator produces an outpainted image of the original size of 192x192, filling the masked part with its hallucinations.
4. Discriminator takes the original ground truth image as input and the outpainted image generated by the generator and tries to predict the probability of how much each of the image is fake.
5. These predictions are utilized for calculating the losses and updating the weights of Generator and Discriminator through backpropagation during the training phase.
6. Once training is completed, testing will done to evaluate the model's capability to outpaint on new images.

Chapter 4

IMPLEMENTATION

In this project, two loss functions are used to evaluate performance of the proposed model. They are Mean Squared Error and L1 loss. These metrics turn out to be the best choice to evaluate the performance of the outpainting model. As mentioned earlier, the training is done using the Landscapes dataset which consists of around 4300 images. Due to computational constraints, the number of images have been reduced to around 1300. Taking into account the size of dataset and time taken for computation the number of training epochs is limited to 150. The architecture uses two Adam optimizers and a constant learning rate of alpha=0.0003 and with beta1=0.5, beta2=0.999 while training the model. To punish the generator for generating bad images as there is a progress in time, the model uses a loss function that varies with time.

During training, in the initial phases the images are cropped to a threshold dimensions set for the model to maintain the uniformity before feeding them into the generator. Initially the images are resized into dimension 192x192, the generator is tasked with the expanding the cropped image of dimension 128x128 back to a 192x192 image.

The generator outputs an outpainted version of a partially masked 192x192 color image, where the masked part in the image is replaced by the model's hallucinations. The generator is equipped with a five layered dense block in which the features from one layer is concatenated and passed to all successive layers, so that they can be further used for accurate down sampling. The intermediate output from the dense block will be passed as an input to subsequent convolutional layers through transition layers in between to optimize the size of the output tensors. The downsampling layers which are subsequent to the upsampling layers helps the generator the get the predicted image to the original size.

The model uses the concept of dual discriminators instead of single discriminator. One of the discriminator is Global Discriminator, that takes the full outpainted image from generator as input and passes it through a series of convolutional layers to classify whether the image is real or outpainted. The second is a Local Discriminator and is made to focus only to the gap between the input image and the outpainted part of the image. Mean of both the discriminators is considered for further layers. The discriminators which does the mean of both the outputs is known as context discriminator. As per discriminators decision the generator will re-organise its weights and generate an another image in the next epoch, which could be better than the previously generated image.

Since the problem of image outpainting has not been widely explored yet and also humans still do a better job at this problem, we end our assessment by providing a subjective discussion about the quality of the outpainted images. Also, since many of the previous proposed models for this problem have very clear noise or irrelevant extrapolations in the outpainted images, we try to evaluate our model manually whether our model is able to overcome this issue or still suffers from similar problems.

Chapter 5

RESULTS AND DISCUSSION

In the below table (Table 5.1), a comparison of the losses for the model for various epochs. Since the loss values don't entirely convey the details of the obtained results, further qualitative discussion on results, and pros and cons in the proposed architecture is presented.

Table 5.1: Comparison of loss results of our model with increase in number of epochs.

	Training			**Validation**		
EPOCH	**L1**	**Adversial**	**MSE**	**L1**	**Adversial**	**MSE**
50	0.0498	0.3584	0.2176	0.0702	0.2725	0.2033
100	0.0389	0.3165	0.2315	0.0712	0.1826	0.2669
150	0.0333	0.3189	0.2298	0.0663	0.4209	0.2417

The observation from the above table (Table 5.1), is that the MSE loss value first increases between 50 to 100 epochs. This happens because the discriminator is out--performing the generator, hence the generator is struggling to generate realistic images that can fool the discriminator. So the generator is punished by increasing the adversarial loss weight relative to L1 loss so that the generator performs better as time progresses.

Table 5.2: Comparison of various models performances.

	Training			**Validation**		
Model	**L1**	**Adversial**	**MSE**	**L1**	**Adversial**	**MSE**
Base Model	0.0601	0.4387	0.1929	0.0846	0.2855	0.2199
Residual Network Model	0.0440	0.4372	0.3256	0.0770	0.4217	0.3511
Proposed Model	0.0333	0.3189	0.2298	0.0663	0.4209	0.2417

From the above tables, all the three loss values are compared. It is observed that the proposed model loss values are significantly lower compared to loss values from majority of the existing architectures mentioned in Table 5.2 for image outpainting. Hence, the proposed architecture outperforms all of the current models. The reason

behind this is the usage of Dense blocks in the Generator which is able to efficiently understand the complex features in the input images and was able to generate much realistic outpaintings.

The Mean Squared Error (MSE) loss is used to measure how well the proposed model can rebuild our ground truth image. But this loss might not be an ideal choice to be used as a quantitative measure for calculating our model quality. The two discriminators, local and global, have different purposes which results is higher MSE.

Outpainted Image

Original Image

Figure 5.1: An example of anomaly within the outpainted images, generated during the training process.

The above figure 5.1 illustrates an anomaly in the proposed outpaintings model. In the outpainted image on the left, the human at the bottom left of the image is distorted and very hard to recognize when compared to the original image on the right, where the person can be found clearly. This tells that the masked image might not contain human which is present in original image when it is provided to the model, thereby the model might think that a human was not present in the image and proceeded to outpaint using this knowledge. This is one of the drawbacks in the proposed model as this might be important.

Finally, the use of Dense blocks have demonstrated the ability of the model to increase image quality in the outpainted parts of the resultant image. The dense blocks along with the transition block seems to increase the ability of the model to generate better image extrapolations, especially at the boundaries of the image, where there is no much context for the model to hallucinate. To outpaint better quality hallucinations, more research into the use of several residual blocks combined with transition blocks can be suggested.

Figure 5.2: Outpainting examples on the Landscapes dataset

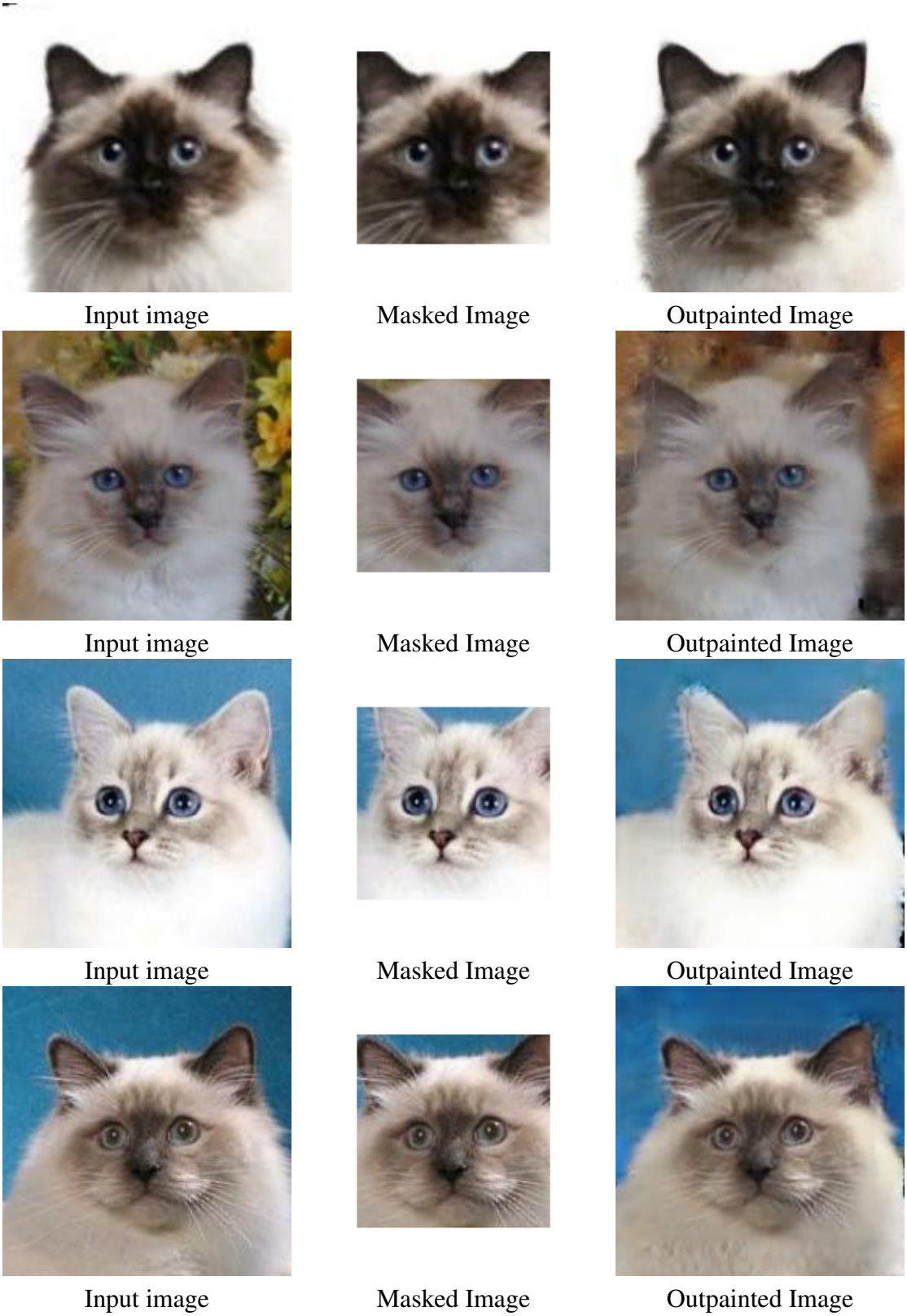

Figure 5.3: Outpainting examples on the cat2dog dataset

Chapter 6

CONCLUSION

Outpainting photos is a useful and challenging process, but it's currently limited by substandard or obviously distorted results. The proposed model using a dense block in the generator combining with a context discriminator results in producing more consistent output images, as well as displays the ability to carry object boundaries into newly created areas. This is an improvement over previous models, and it showed promising results.

Chapter 7

FUTURE ENHANCEMENT

Currently, outpainting is limited to images only. So it can be future expanded to make the model more suitable for videos. This can be achieved by extracting all the frames of the videos as images and each image is outpainted. These outpainted images(frames) are combined back to form an outpainted video.

The biggest challenge that was encountered while conducting the experiment was the training set size and the number of epochs we could train our model. Since the number of images is reduced drastically from 4300 to 1300, the model was not able to effectively learn complex features such as human, water and tree details from an image. This problem further gets stacked since there were a very small subset of images that had humans in them. So in the future, the model can be trained on the full length of dataset so that the above problem could be resolved.

As there is a fixed average between local and global discriminator in the proposed model, it makes the model to weigh the overall image and boundaries equally, leads to the output images generated to be of lower quality(blurred). The probable way to overcome this issue is to form a function that back propagates to discriminator which will allow the model itself to decide on how to mix the outputs from local and global discriminator.

REFERENCES

1. Gardias, P., Arthur, E., and Sun, H. (2020). "Enhanced residual networks for context-based image outpainting.

2. Kyunghun Kim, Yeohun Yun, K.-W. K. K. K. S. L. and Kang, S.-J. (2020). "Painting outside as inside: Edge guided image outpainting via bidirectional rearrangement with progressive step learning.

3. Sabini, M. and Rusak, G. (2018). "Painting outside the box: Image outpainting with gans.

4. Wang, Y., Tao, X., Shen, X., and Jia, J. (2019). "Wide-context semantic image extrapolation." *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1399–1408.

5. Xiao, Q., Li, G., and Chen, Q. (2020). "Image outpainting: Hallucinating beyond the image." *IEEE Access*, 8, 173576–173583.

6. Xiaofeng Zhang, Feng Chen, C. W. S. W. M. T. G. J. (2020). "Sienet: Siamese expansion network for image extrapolation.

7. Zhang, L., Wang, J., and Shi, J. (2019). "Multimodal image outpainting with regularized normalized diversification." *CoRR*, abs/1910.11481.

8. Zongxin Yang, Jian Dong, P. L. Y. Y. S. Y. (2019). "Very long natural scenery image prediction by outpainting.