SEGMENTATION OF BREAST LESIONS FROM ULTRASOUND IMAGES USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

A thesis submitted to the
Department of Electrical & Electronic Engineering
of
Bangladesh University of Engineering & Technology

In partial fulfillment of the requirement for the degree of
Master of Science in Electrical & Electronic Engineering

Submitted by

SATYAJEET DATTA
Student Id: 0416062216

Department of Electrical & Electronic Engineering (EEE)
BANGLADESH UNIVERSITY OF ENGINEERING & TECHNOLOGY (BUET)

MARCH 2022

The thesis titled **"SEGMENTATION OF BREAST LESIONS FROM ULTRA-SOUND IMAGES USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK"**, submitted by SATYAJEET DATTA, Student Id: 0416062216, Session: April 2016, has been accepted as satisfactory in partial fulfilment of the requirement for the degree of Master of Science in Electrical & Electronic Engineering on March 12, 2022.

BOARD OF EXAMINERS

12/3/2022

Dr. Mohammad Ariful Haque,
Professor,
Department of Electrical & Electronic Engineering,
BUET, Dhaka- 1205.

Chairman
(Supervisor)

Dr. Md. Kamrul Hasan,
Professor and Head of the Department,
Department of Electrical & Electronic Engineering,
BUET, Dhaka-1205.

Member
(Ex-Officio)

Dr. Mohammed Imamul Hassan Bhuiyan,
Professor,
Department of Electrical & Electronic Engineering,
BUET, Dhaka-1205.

Member

Dr. Mohammad Rakibul Islam,
Professor,
Department of Electrical & Electronic Engineering,
Islamic University of Technology,
Board Bazar, Gazipur- 1704, Bangladesh.

Member
(External)

CANDIDATE'S DECLARATION

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Satyajeet

SATYAJEET DATTA

ACKNOWLEDGMENT

I would love to express my heartiest gratitude to my supervisor Dr. Mohammad Ariful Haque, Professor, Department of Electrical and Electronic and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka. His supervision, encouragement and personal guidance were always with me during the progress of the thesis. His concepts and problem solving approach have been very helpful for the successful completion of this work. I am ever grateful to him.

Finally, and most importantly, I remember my parents and family members, specially my mother, who are always a source of inspiration in every good action. I am grateful to them for their faith in me and allowing me to be as ambitious as I wanted. It was under their watchful eye that I gained so much drive and an ability to tackle challenges ahead on.

Dedication

This thesis is dedicated to my late father who brought uniqueness in my thinking process, to my beloved mother who always provides me mental strength to combat with challenges and last of all to all my family members who provided me mental support.

Contents

Page

CANDIDATE'S DECLARATION . . . iii

ACKNOWLEDGMENT . . . iv

DEDICATION . . . v

TABLE OF CONTENTS . . . vi

LIST OF TABLES . . . ix

LIST OF FIGURES . . . x

LIST OF ACRONYMS . . . xii

ABSTRACT . . . xiv

CHAPTER . . . 1

1 Introduction . . . 1
- 1.1 Motivation . . . 1
- 1.2 Challenges associated with the research . . . 3
- 1.3 Segmentation basic . . . 4
- 1.4 Contribution of the thesis . . . 5
- 1.5 Outline of the thesis . . . 6

2 Introduction to Generative Adversarial Network and Literature review 8
- 2.1 Popular Deep Learning Architectures . . . 9
- 2.2 Generative vs Discriminative models . . . 11
- 2.3 Different types of generative networks: . . . 11
  - 2.3.1 Autoencoder . . . 11
  - 2.3.2 Generative Adversarial Network: . . . 13

2.4 Training of GANs . . . 14
2.5 Applications of GAN . . . 15
2.6 Common Challenges for GAN models . . . 16
2.7 Recent practice to mitigate problems in training GAN . . . 17
2.7.1 Loss assessment from the intermediate layer . . . 18
2.7.2 Optimizing the model by mini-batch discrimination . . . 19
2.7.3 Avoiding over-fitting by historical averaging . . . 21
2.7.4 One-sided label smoothing . . . 21
2.7.5 Using labels (Conditional GAN) . . . 21
2.7.6 Cost functions . . . 22
2.8 Insight about CGAN . . . 24
2.8.1 Basic Architecture . . . 25
2.8.2 Methodology . . . 26

3 Biomedical Image Segmentation via Deep Learning . . . 29
3.1 Evolution of Deep Learning in Medical Imaging . . . 29
3.2 Deep Learning in Medical Imaging Applications . . . 31
3.3 Medical Image Segmentation via deep learning . . . 32
3.3.1 Generative Adversarial Network in Medical Imaging Analysis . . . 35
3.4 Cases for Ultrasound Images . . . 36

4 Proposed Segmentation Method: Modified-CGAN (Mod-CGAN) . . . 39
4.1 Multi-Task Learning . . . 39
4.2 Proposed Method: Modified - CGAN (Mod-CGAN) . . . 41
4.2.1 Basic Architecture . . . 43
4.2.2 Training Structure . . . 44
4.2.3 Test Structure . . . 47
4.3 Features of Mod-CGAN . . . 48
4.4 Objective Function . . . 49
4.5 Distinction with Multi Tasking based GAN architecture . . . 51

5 Performance Evaluation of the Proposed System . . . 54
5.1 Dataset Details . . . 55
5.2 Simulation Environment . . . 55
5.3 Analysis . . . 57
5.3.1 Quantitative Assessment of Segmentation performance . . . 57
5.3.2 Loss Function Analysis . . . 58

5.3.3 Different versions of Mod-CGAN 60
5.3.4 Performance Evaluation for Dataset II 66
5.4 Performance evaluation of Modifier imposition for different CGAN architecture 67
5.5 Qualitative interpretation on Dataset I 69
6 Conclusion 71
REFERENCES 74

List of Tables

1.1 Mathematical Definition of different types of Evaluation Metric . . 5
2.1 Cost functions assoiated with different types of GANs 23
2.2 FID score associated with different types of GANs for different datasets from [41] 24
5.1 Structural Difference between Pix2Pix and Mod-CGAN 56
5.2 Quantitative Assessment on the dataset 57
5.3 DSC associated with different variants of Mod-CGAN 62
5.4 Quantitative Assessment on Dataset II 67
5.5 Quantitative Assessment between Pix2Pix and SegNet-cGAN . . . 67
5.6 Structural Features of SegNet-cGAN and Mod-SegNet-cGAN . . . 68
5.7 Quantitative Assessment on the dataset 68

List of Figures

1.1 Segmentation map for medical image 4
2.1 Graphical representation of an autoencoder network 12
2.2 Block diagram of GAN architecture 13
2.3 Illustration of mode collapse on MNIST dataset from [38] 17
2.4 Full framework for Mini-batch discrimination 20
2.5 Discriminator framework in Mini-batch discrimination 21
2.6 Framework for Conditional GAN 22
2.7 Different parts of conditional generative adversarial network 25
2.8 Example results on Google Maps at 512x512 resolution 27
2.9 Example results on Cityscape Labels → photo, compared to ground truth 27
2.10 Example results of our method on facades labels → photo, compared to ground truth be implemented via Pix2Pix framework along with our proposed upgradation methodology to Pix2Pix framework and also it will be shown how both the methods surpass state of the art result in the field of medical imaging segmentation task. 28
3.1 U-net architecture 34
3.2 SegNet-cGAN Architecture 37
3.3 Illustration of Medical US Analysis 38
4.1 single-input multi-output (SIMO), where single set of input are mapped to multiple target 41
4.2 Basic architecture of Mod-CGAN 43
4.3 Detailed Modifier Architecture 44
4.4 Detailed Generator Architecture 45
4.5 Detailed Discriminator Architecture 46
4.6 Architecture during test time 47
4.7 Diagram for Generator of CGAN 48

4.8 **Diagram for Mod - CGAN** 49
4.9 **Multi Tasking GAN from [72]** 52
4.10 **Diagram for MTGAN from [73]** 52
4.11 **Task based Mod-CGAN implementation** 53
5.1 **Loss functions regarding three networks of Mod-CGAN** 58
5.2 **Loss functions regarding two networks of CGAN** 59
5.3 **Generator loss functions for two models: CGAN and Mod-CGAN** 60
5.4 **Discriminator loss functions for two models: CGAN and Mod-CGAN** 61
5.5 **Depiction for different versions of Mod-CGAN** 63
5.6 **Generator loss function (Original and Zoom Version) for different versions of Mod-CGAN** 64
5.7 **Discriminator loss function (Original and Zoom Version) for different versions of Mod-CGAN** 65
5.8 **Modifier loss function for different versions of Mod-CGAN** 66
5.9 **Qualitative Assessment on dataset I** 70

LIST OF ACRONYMS

ACGAN Auxiliary Classifier GAN

AI Artificial Intelligence

BEGAN Boundary Equilibrium GAN

CGAN Conditional GAN

CAD Computer Aided Diagnosis

DM Digital Mammography

DRAGAN Deep Regret Analytic GAN

DSC Dice Score

EBGAN Energy-based GAN

FCN Fully Convolutional Network

FID Frechet Inception Distance

GAN Generative Adversarial Network

infoGAN Information Maximizing GAN

LSGAN Least Squares GAN

MMGAN Manifold Matching GAN

MTL Multi- Task Learning

MTGAN Multi Tasking GAN

NSGAN Non- saturating GAN

US Ultrasound

VAE Variational Auto Encoder

WGAN Wasserstein GAN

WGAN-GP Wasserstein GAN + Gradient Penalty

Abstract

In the way to facilitate the scope of Computer Aided Diagnosis (CAD) into the treatment of breast cancer, which is a leading issue of concerns for women worldwide in recent times, the task of breast lesion segmentation is a very critical processing step that needs to be automated. Although Digital Mammography (DM) is the most popular screening tool in breast cancer detection, Ultrasound (US) imaging has recently emerged as a popular alternative due to its non-invasive nature, real time and low cost imaging.

Breast lesion segmentation from US images using deep learning techniques is quite challenging. US images contain many fuzzy contours and false edges along with the original mask. Again, there has been shortage of publicly available large annotated datasets of Breast US images for training the deep learning model. Moreover, the introduction of adversarial training for segmentation task has been quite nascent which poses major challenges of convergence and stability issues.

We have implemented a Conditional Generative Adversarial Network (CGAN) based approach for the task of breast lesion segmentation from US Images. Specifically, the network has been designed as an upgradation to the architecture associated with CGAN by imposing multi tasking learning in the training process. Convergence as well as stability of the newly designed model has been largely improved compared with CGAN. Also, overall performance of the segmentation task has been assessed in terms of the state of the art model such as U-Net, Pix2Pix, SegNet-cGAN. In addition to this, performance improvement has been attained for different scenarios such as different dataset, different model etc.

Chapter One

Introduction

1.1 Motivation

The rapid advancement of deep learning approaches in various practical fields continues to fuel the medical imaging community's interest in implementing these techniques to improve the accuracy of cancer screening. Breast cancer has been a leading issue of concern for women worldwide in recent times. As mentioned in [1], more than 8% of women will develop breast cancer during their lifetime. However, periodic clinical checkups and self-tests help in early detection and thereby significantly increase the scope of survival. Early detection of breast cancer as well as accurate assessment of lesions are the goals of various imaging modalities. The most commonly used screening tool for this purpose is the Digital Mammography [2]. Mammography is a specialized medical imaging tool that uses a low dose X-Ray system to detect cancer early before women experience symptoms and when it is most treatable. Digital Mammography (DM), also called full-field digital mammography (FFDM), is a mammography system in which the x-ray film is replaced by electronics that convert x-rays into mammographic pictures of the breast. These systems are similar to those found in digital cameras and they are able to produce better pictures with a lower radiation dose. These images of the breast are transferred to a computer for review by the radiologist and for long term storage. While mammography is the most widely used screening tool for breast cancer, mammograms do not detect all breast cancers [3]. This is called a false negative result. On the other hand, when a mammogram looks abnormal but no cancer lesion is present, this is called a false positive result. Therefore, in order to determine the existence of a benign or malignant tumor, only one screening tool is not enough. Again, interpretations of mammograms can be difficult because a normal breast looks different for each woman. In addition to this, the appearance of an image may be compromised if there is powder or salve on the breasts or if the examined patient has undergone breast surgery. Recently increased breast density has attracted attention due to the limitations of mammography in terms of finding cancer[4][5]. Apart from those described above, breast implants can also impede accurate mammogram readings because both silicone and saline implants are not transparent on x-rays and can block a clear view on tissues around them, especially if the implant has been placed in front of, rather than beneath, the chest muscles. In terms of risks, there is always a slight chance of cancer from excessive exposure to radiation.

Considering all these serious limitations in DM, research has been ongoing on a variety of breast imaging techniques that can contribute to the early detection of breast cancer and improve the accuracy in distinguishing non-cancerous breast conditions from breast cancer. Recently, Ultrasound (US) Imaging is being considered as an alternative to DM imaging due to leading a very important role in breast cancer detection [6], image guided biopsy [7] and lymph node diagnosis. US imaging of the breast uses sound waves to produce pictures of the internal structures of the breast. Because US images are captured in real-time, they can show the structure and movement of the body's internal organs. It is safe, noninvasive, widely available, easy-to-use, less expensive than most other imaging methods and most of all, does not use radiation. Its performance on the weaker aspects of DM imaging methods is quite satisfactory. Still it poses the limitations of poor quality images caused by speckle noise, low contrast and shadow effect [8]. Considering the non-invasive nature of US imaging, a sophisticated Computer Aided Diagnosis (CAD) model can be designed to assist the radiologists in detecting and segmenting the breast lesions more efficiently and effectively.

1.2 Challenges associated with the research

Breast US images contain too much confusing contours due to noise associated with it. So it is very difficult to find out the region of interest regarding breast lesions. It is quite obvious that the path of tracing the region of interest from the breast US images among the fuzzy, misleading surrounding region is not that much easy to perform. Along with it, the scarcity and unavailability of largely annotated datasets complicates the problem in US based breast lesion segmentation.

The popular deep learning techniques applied to image segmentation are U-Net segmentation [9], multiple domain features [10], Patch based Le-Net [6], a transfer learning approach with a Pre-trained FCN- AlexNet [7] etc. However, these approaches have been practiced much on datasets of very large size which is very usual practice for computer vision based tasks. Recently, learning from adversaries has gained huge reputation in solving many computer vision based tasks on natural images such as Image Synthesis, Image to Image Translation [7], Superresolution [11] etc. There is a trend to replace all discriminative approaches to solve different computer vision based problems with different versions of Generative Adversarial Network (GAN) in a way to make the network more smart and immune to external adversaries. Although due to the robustness in the training procedure [12], the training of GAN imposes major difficulties due to instabilities caused by a difficult minimax optimization problem leading to mode collapse both partially and completely as will be discussed in the subsequent chapter. Again, as breast US images contain too much confusing contours due to noise associated with it and also GAN has its implementation complexities associated with the design of generator and discriminator loss functions subject to more insightful research, GAN architecture has not been practiced in breast US segmentation task yet.

1.3 Segmentation basic

Medical image segmentation is a vital area in medical image analysis and is necessary for diagnosis, monitoring and treatment. Due to the variable size, shape and location of the target tissue, medical image segmentation is one of the most challenging tasks in medical image analysis. The main purpose of the segmentation task is to assign label to each pixel of the images. It normally includes two steps: firstly, detect the unhealthy tissue or areas of interest; secondly, delineate the different anatomical structures or areas of interest. Figure 1.1 depicts the summary of overall segmentation task.

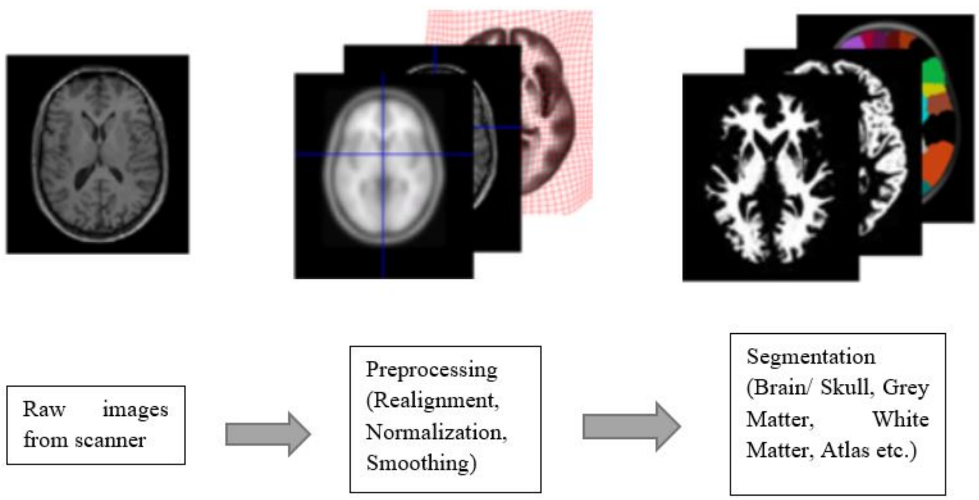


Figure 1.1 Segmentation map for medical image

Despite the variety of the proposed segmentation network architectures, it is still hard to compare the performance of different available architectures. In order to obtain accurate segmentation and compare different state of the art methods on the same ground, some well-known public challenges for segmentation have been developed, such as Brain tumor Segmentation (BraTS) [13], Ischemic Stroke Lesion Segmentation (ISLES), Neonatal Brain Segmentation [14], MR Brain Image Segmentation (MRBrainS) [15], Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS), 6-month infant brain MRI Segmentation

(Iseg-2017) [16] and Automatic intervertebral disc localization and segmentation from 3D Multi-Modality MR (M3) Images (IVDM3Seg). Table 1.1 depicts some commonly used evaluation metrics in performing the segmentation task.

Evaluation Metric	**Mathematical Description**
Dice Score (DSC)	$DSC = \frac{2*TP}{2*TP+FP+FN}$
Sensitivity	$Sensitivity = \frac{TP}{TP+FN}$
Specificity	$Specificity = \frac{TN}{TN+FP}$
Hausdorff distance (HD)	$HD = max\{sup_{r \in dR} d_m(s,r),$ $sup_{s \in dS} d_m(s,r)\}$
Average boundary distance (ABD)	$ABD(X_s, Y_s) = \frac{1}{N*X_s+N*Y_s}(\sum_{x \in X_s} min_{y \in Y_s} d(x,y)$ $+ \sum_{y \in Y_s} min_{x \in X_s} d(y,x))$

Table 1.1 Mathematical Definition of different types of Evaluation Metric

Here, TP, FP, FN and TN refer to True Positive, False Positive, False Negative and True Negative pixels respectively. For the case of Hausdorff distance (HD), dS and dR are the sets of lesion border pixels/voxels for the predicted and the truth segmentations, and $d_m(v,u)$ is the minimum of the Euclidean distances between a voxel v and voxels in a set u. For ABD, X_s and Y_s are the sets of surface points of the reference and algorithm segmentations respectively. The operator d is the Euclidean distance operator.

1.4 Contribution of the thesis

This thesis deals with the complex task of US image segmentation with a view to making the CAD system more precise by adopting an adversarial training procedure. Our contribution of the thesis are as follows:

- We have developed a deep learning method based on adversarial training so that the GAN model can be utilized in the biomedical image segmentation task. It can also be mentioned that GAN is replacing many discriminative deep learning models with better generalization performance.

- A new conditional GAN model termed as Modified CGAN (Mod-CGAN) has been developed to utilize the labeled data more efficiently. The training stage has been divided into two steps such that it takes much less iteration to converge as compared to the conventional conditional GAN model. The proposed model can also perform quite satisfactorily without utilizing any data augmentation task.
- The way our proposed method has been implemented, it can be assimilated to the concept of self supervised learning. For the purpose of learning in the extra added network named modifier, no further annotation is required. Available raw image as well as ground truth image will suffice to the learning of the modifier network.
- The experimental result shows that robustness of the GAN framework can be improved by training both generator and discriminator for an extended number of epochs even after the equilibrium.
- In terms of performance, our proposed method surpasses the state of the art model U-Net as well as the raw Conditional GAN model by a moderate margin.

1.5 Outline of the thesis

The subsequent chapter has been designed as follows:

- Chapter 2 discusses briefly about the emergence of deep learning framework in solving complex problems. Then, it mentions some of the most popular deep learning architectures established to replicate the human level performance. It introduces the reader with the difference between discriminative and generative architectures and henceforth, the frameworks associated with generative adversarial network have been discussed. Also, it includes some problems associated with the training of GAN. In the last part of the chapter, it specifies some recent practice in mitigating the convergence and mode collapsing problem.

- Chapter 3 mentions about the evolution of deep learning in biomedical imaging applications and also briefly discusses some of the fields where deep learning has been incorporated. Then, it broadly discusses specifically about biomedical image segmentation task via deep learning. After that, a brief overview about GAN in medical imaging analysis has been included. In the end, some architectures regarding biomedical image segmentation task from the literature have been covered.
- Chapter 4 discusses the proposed methodology, objective function of our framework as well as the intuition behind the expected improvement. Explicitly it also covers the detailed architecture associated with different parts of our proposed method.
- Chapter 5 is actually about discussing and analyzing the performance of the proposed method. Our method along with two other familiar methods (one is state of the art segmentation approach named U- Net and other is the basic Conditional GAN approach) have actually been simulated to make the comparison more transparent and fair. Again, those models have been evaluated not only qualitatively but also quantitatively. Also there have been covered different versions of our proposed method as well as the loss function analysis associated with the model.
- Chapter 6 mentions about the future scope associated with this proposed method and also recommends to incorporate it in many computer vision tasks of natural image in a way to mitigate the problems with basic framework of GAN.

Chapter Two

Introduction to Generative Adversarial Network and Literature review

People often throw around words like "artificial intelligence" and "neural networks" and "machine learning" and "deep learning" quite interchangeably. But significance of these words are not similar. The history of Artificial Intelligence (AI) predated to 1956 in United States, where engineers decided they would write a computer program that would try to imitate natural intelligence. Within AI, a new field emerged called machine learning. Instead of writing a step-by-step program to do something - which is a traditional approach in AI — people may collect lots of data about something that they are trying to understand. For example, they are trying to recognize objects, so they collect lots of images of their interest. Then, with machine learning, it's an automated process that dissects out various features, and figures out that one thing is an automobile and the other is a stapler. It is a very large field and goes way back. Originally, people were calling it "pattern recognition," but the algorithms became much broader and much more sophisticated mathematically. Within machine learning are neural networks inspired by the brain. In the mid-1980s and early 1990s, many important architectural advancements were made in neural networks. However, the amount of time and data needed to get good results slowed adoption, and thus interest cooled. In the early 2000s, computational power expanded exponentially and the industry saw a "Cambrian explosion" of computational techniques that were not possible prior to this. Deep learning emerged from that decade's explosive computational growth as a serious contender in the field, winning many important machine learning competitions. The contribution from [17], [18], [19], [20] [21] heavily ensures the acceleration in the current deep learning frameworks. However, the sheer dominance of deep learning in the world can be pinpointed to the particular moment of history: December 2012 at the NIPS meeting, which is the biggest AI conference. There, computer scientist Geoff Hinton and two of his graduate students showed it is feasible to take into account a very large dataset called ImageNet, with 10,000 categories and 10 million images, and reduce the classification error by 20 percent using deep learning. With the availability of more data and complex data structure, here emerges the introduction of deep learning. Deep learning is a branch of machine learning that deploys algorithms for data processing, decision making and automatic feature extraction in a way to imitate the thinking process and even develop abstractions. Deep networks are trained from the data in the same way that babies experience the fresh world, adopts with the ongoing world and thereby gradually acquires the skills needed to navigate novel environments. Learning algorithms extract information from raw data; information can be used to create knowledge; knowledge underlies understanding and thereby leading to wisdom. In the next subsequent section, a brief insight about some popular deep learning architectures will be covered. From then, we will focus our attention to generative models and thereby will have a thorough description about Generative Adversarial Network.

2.1 Popular Deep Learning Architectures

Google's inception network is an advanced and deep architecture that was applied successfully for several tasks [22] like classification, detection etc. Its main highlight is the introduction of the so-called inception block that essentially allows to compute convolutions and pooling operations in parallel. By repeating this block in a network, the network can select by itself in which sequence convolution and pooling layers should be combined in order to solve the task at hand effectively.

ResNets have been designed to enable training of very deep networks [23]. Even with earlier methods, networks will not benefit from more than 30 to 50 layers, as the gradient flow becomes numerically unstable in such deep networks. In order to alleviate the problem, a so-called residual block is introduced, and layers take the form $f(x) = x + f_(x)$, where $f_(x)$ contains the actual network layer. Doing so has the advantage that the addition introduces a second parallel branch into the network that lets the gradient flow from end to end. ResNets also have other interesting properties, e.g., their residual blocks behave like ensembles of classifiers [24].

Variational networks enable the conversion of an energy minimization problem into a neural network structure [25]. We consider this type of network as particular interesting, as many problems in traditional medical image processing are expressed as energy minimization problems. The main idea is as follows: energy function is typically minimized by optimization programs such as gradient descent. Thus, we are able to use the gradient of the original problem to construct a so-called variational unit that describes exactly one update step of the optimization program. Succession of such units then describe the complete variational network. Two observations are noteworthy: First, this type of framework allows to learn operators within one variational unit, such as parsifying transform for compressed sensing problems. Second, the variational units generally form residual blocks, and thus variational networks are always ResNets as well.

Recurrent Neural Networks (RNNs) enable the processing of sequences with long term dependencies. Furthermore, recurrent nets introduce state variables that allow the cells to carry memory and essentially model any finite state machine. Extensions are long short-term memory (LSTM) networks [26] and gated recurrent units (GRU) [27] that can model explicit read and write memory transactions similar to a computer.

2.2 Generative vs Discriminative models

Before digging deep into the details of Generative Adversarial Network (GAN), we first have to make ourselves familiar with the idea of generative modelling in contrast to discriminative modelling. Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset. In contrast, discriminative modeling refers to find a discriminant function f(x) that maps each x directly onto a class label, thereby combining the decision and inference stages into a single learning problem. So, generative modeling tends to create or generate new examples in the input distribution after sufficiently summarizing the input distribution. In terms of complexities, generative modeling tackles a more difficult task due to trying to figure out how data is placed throughout the space in contrast to draw the boundaries in the data space for the case of discriminative modeling. So in generative modeling, the challenge is to imitate the input data distribution from an unsupervised training of different types of tasks. The idea behind the recent progress of generative modeling is to convert the generation problem to a prediction one and use deep learning algorithms to learn such a problem.

2.3 Different types of generative networks:

2.3.1 Autoencoder

Autoencoders and their encoder/ decoder frameworks are the inspiration behind generative models. It work by taking input and generating a smaller vector representation for the purpose of later reconstructing its input. It is done by using an encoder to impose an information bottleneck on input data and then utilizing a decoder to recreate the input data based on that representation. This is based on the idea that there are hidden structures within the data (that is, correlations, and so on), but that are not readily apparent. Autoencoders are a means of automatically learning these relationships without explicitly doing so. Structurally, autoencoders consist of an input layer, a hidden layer and an output layer as shown in Figure 2.1. The encoder learns to preserve as much of the relevant information as possible in the limited encoding and intelligently discards irrelevant parts. This forces the network to maintain only the data required to recreate the data; we do this using a reconstruction loss

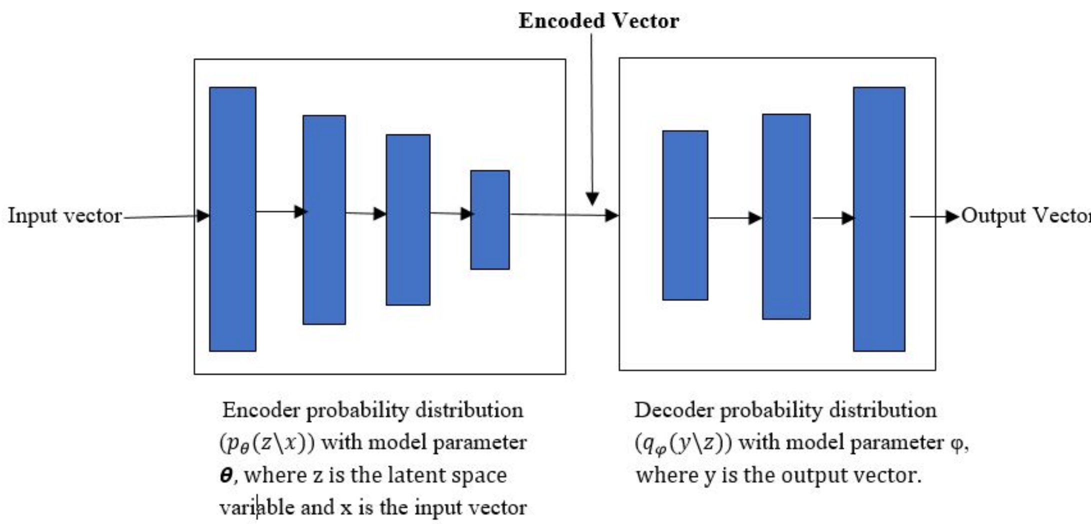


Figure 2.1 Graphical representation of an autoencoder network

with a regularization term to prevent over fitting. It is particularly useful in the task of dimensionality reduction [28] due to its performance in both linear and nonlinear manifolds in contrast to the basic technique of Principal Component Analysis. There have been used different types of autoencoders depending on the task of encoding [29], [30]. Although it offers a much suitable way to deal with compression and inter layer transfer learning, yet it produces blurry samples considering the no. of parameters in the bottleneck representation where in the bottleneck layer, higher no. of parameters can't be allowed. Also, it imposes the risk of not learning the true posterior distribution.

2.3.2 Generative Adversarial Network:

It is revealed in recent works that deep learning models are often vulnerable to adversarial examples [31], [32]. So, adversarial examples are maliciously designed to deceive the target model by generating carefully crafted adversarial perturbations on original clean inputs. In response to this susceptibility of deep learning models, researchers have found the adversarial training as a promising approach [33] to defend against different types of real time adversaries in contrast to the stationary and benign environments assumption for training and test data. In motivation for introducing adversaries in the learning process, a new type of generative modeling approach invented by IAN Godfellow and his colleagues in 2014 [34], has been designed in the way to solve the min-max problem. This network is called Generative Adversarial Network (GAN). GANs are based on a game theoretic scenario with an objective

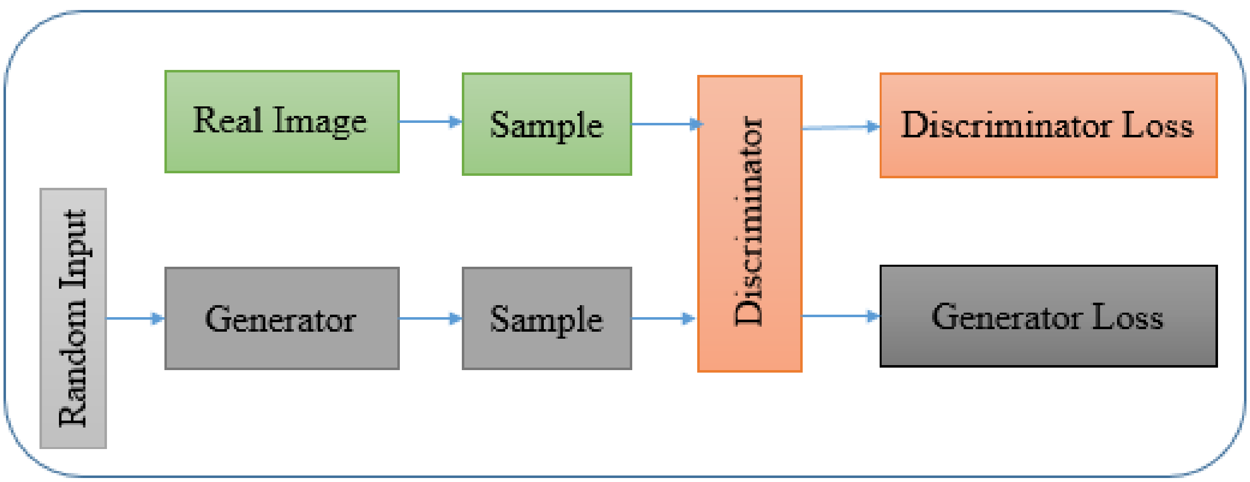


Figure 2.2 Block diagram of GAN architecture

to find Nash equilibrium between two networks, Generator and Discriminator. The idea is to sample from a simple distribution like Gaussian and then learn to transform this noise to data distribution using universal function approximators such as neural networks. The generator network must compete against an adversary. The generator network directly produces samples. Its adversary, the discriminator network, attempts to differentiate between the samples drawn from the generator and from the training data. Figure 2.2 illustrates how GAN works. The generator output is connected directly to the discriminator input. As described earlier, the GANs are formulated as a minimax game, where the Discriminator is trying to minimize its reward, $V(D,G)$, and the Generator is trying to minimize the Discriminator's reward, in other words maximize its loss. It can be mathematically formulated as below:

$$\min_{G} \max_{D} V(D,G)$$

$$V(D,G) = E_{x \sim p_{data}(x)}[logD(x)] + E_{z \sim p_z(z)}[log(1 - D(G(z))] \tag{2.1}$$

Where

$$G = \quad \text{Generator}$$

$$D = Discriminator$$

$$P_{data}(x) = \quad \text{distribution of real data}$$

$$P_z(z) = \quad \text{distribution for generator latent variable}$$

$$x = \quad \text{sample from } P_{data}(x)$$

$$z = \quad \text{sample from } \mathrm{P}_z(z)$$

$$E_{x \sim p_{data}(x)} = \quad \text{Expectation operator acted on sample from the distribution of real data}$$

$$E_{z \sim p_z(z)} = \quad \text{Expectation operator acted on sample from the distribution of generator latent variable}$$

2.4 Training of GANs

Phase 1: The Discriminator is trained while the Generator is idle. In this phase, the network is only forward propagated and no back-propagation is done. The Discriminator is trained on real data to have the idea of real data distribution, and see if it can correctly predict them as real. Also, in this phase, the Discriminator is also trained on the fake generated data from the Generator and see if it can correctly predict them as fake. Mathematically, the discriminator network can be formulated as follows:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} [logD(x^{(i)}) + log(1 - D(G(z^{(i)})))]$$

Here, m is the number of samples in a batch, θ is the model parameter in the generic sense where subscript "g" signifies about the generator operation. The first part within the summation makes sure that real samples are classified as being real and the second part makes sure generated samples are classified as unreal. Discriminator, D wants to maximize the first part and also intends to minimize the second part.

Phase 2: The Generator is trained while the Discriminator is idle. After the Discriminator is trained by the generated fake data of the Generator, we can get its predictions and use the results for training the Generator and get better from the previous state to try and fool the Discriminator. Mathematically, the generator network can be formulated as follows:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} [log(1 - D(G(z^{(i)})))]$$

Here, subscript "d" signifies about the discriminator operation. Generator, G determines how realistic the generated samples are. G wants to maximize this. In this way, after completing reasonable number of iterations/ epochs, convergence will be done after ensuring the fact that the generator is clever enough to mislead the discriminator. In the space of arbitrary functions G and D, a unique solution exists.

2.5 Applications of GAN

After the invention of GAN, it has been widely used to replace normal convolutional neural network based deep learning approaches in many computer vision tasks. And today, it has already outclassed many state of the art results based on normal deep learning approach. Some of its applications has been listed below:

- **Image Synthesis**: This includes font generation, Text2Image as well as 3D image generation [35].
- **Data Augmentation:** It aims to reduce the need for labeled data (GAN is only used as a tool for enhancing the training process of another model). Because of this, it is possible to sample from generator which we can use to enhance our original dataset.
- **Style Transfer and Manipulation:** It includes Face Aging, painting, pose estimation and manipulation, inpainting, and blending [36].
- **Image Super-Resolution:** Image super resolution means to increase the resolution of the image artificially from the available low resolution images. GAN supports quite satisfactorily in this ill-posed case [37] as there are always available many High Resolution images corresponding to Low Resolution images.

2.6 Common Challenges for GAN models

- **Mode Collapse and mode drop:** One of the major failure modes in training a GAN model is mode collapse both partially and fully. Mode collapse is a problem that refers to the situation when the generator generates samples having little variety or the generator generates same samples at every steps. In terms of probability distributions, sometimes the probability distribution is multi-modal and very complex in nature, means that it contains samples from different observations. In this case, GANs fail to model the multi-modal distributions of data and suffers from mode collapse. As depicted from [38] in Figure 2.3, the two rows are produced by two different GANs, where top row produce digits from all modes (from 0 to 9), but bottom row only produces images of mode 6 (digit 6).
- **Non Convergence:** According to game theory, the GAN model converges when the

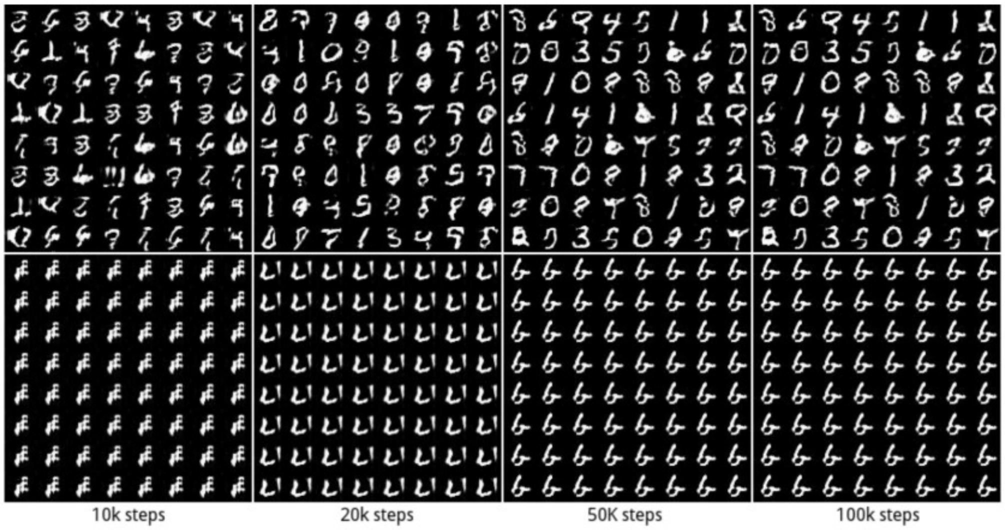


Figure 2.3 Illustration of mode collapse on MNIST dataset from [38]

generator and the discriminator reach a Nash equilibrium. Since the actions of both sides are fully in reverse direction, a Nash equilibrium happens when one side does not change its action regardless of what the opponent will do. This can be referred to the fact that Gradient descent based GAN optimization does not correspond to convex-concave game even for simple parameterizations [38]. The generator objective function of the first original GAN paper has the issue of vanishing gradients and the alternative cost function has fluctuating gradients that cause instability to the training of the models [39]. Moreover, as mentioned in [39], if the supports of unknown data distribution and generator data distribution are disjoint or lie in lower dimensional manifolds, then there is always a perfect discriminator between them leading to an unreliable training of the generator.

2.7 Recent practice to mitigate problems in training GAN

In this section, we will try to have some insight on how to improve GAN. In practice, the following practices have been exercised in the recent literature in order to mitigate the problems of GAN.

- Penalize the generator and discriminator based on the loss functions assessed from the intermediate layer
- Change the cost function for better optimization goal
- Add additional penalties to the cost function to enforce constraints
- Avoid overconfidence and overfitting
- Better ways of optimizing the goal
- Add labels

As research regarding optimization with GAN has been a dynamic topic, so quest for new techniques has been always ongoing to adapt this architecture in solving many complex problems. Now, we will some of the above mentioned techniques in more details.

2.7.1 Loss assessment from the intermediate layer

During the training of GANs, we maximize the objective function of the discriminator network and minimize the objective function for the generator network. But there has been some serious flaws in the training process as it does not count on the statistics of real data and generated data. Feature matching technique was proposed by Tim Salimans, Ian Goodfellow and others in their paper [40] to improve the convergence of GANs by introducing a new objective function, thereby rectifying some flaws in the training process. The new objective function leans the generator to generate data with statistics that is similar to the real data. Here, the network does not prompt the discriminator to provide binary labels, instead, the discriminator provides activations or feature maps of the input data, extracted from an intermediate layer in the discriminator network. From the training perspective, we train the discriminator network to learn the important statistics of the real data. So, discriminator learns those discriminative features. The new objective function can be described mathematically as follows. Let's assume f(x) – the activation or feature maps for the real data from an intermediate layer in the discriminator network;

f(G(z)) – the activation or feature maps for the data generated by the generator network from an intermediate layer in the discriminator network.

Now we can represent the new objective function as:

$$||E_{x \sim p_{data}} f(x) - E_{z \sim p_z(z)}) f(G(z))||_2^2$$

This objective function can achieve better results compared to the conventional objective function mentioned in equation (2.1).

2.7.2 Optimizing the model by mini-batch discrimination

During the phase of mode collapse, all images created look similar. To mitigate the problem, real images and generated images can be fed into the discriminator separately in different batches and compute the similarity of the image x with images in the same batch. If the mode starts to collapse, then the similarity of the generated images increases. The overall similarity $O(x_i)$ between the image x_i and other images in the same batch is computed after passing through a Transformation matrix T, which is of dimension A X B X C.

In Figure 2.4, x_i is the input image and x_j is the rest of the images in the same batch. Transformation matrix is to transform the features x_i to M_i which is a B X C matrix. So, the individual similarity $c(x_i, x_j)$ can be derived between image x_i and x_j using the L_1-norm.

$$c_b(x_i, x_j) = exp(-||M_{i,b} - M_{j,b}||_1)$$

The overall similarity $O(x_i)$ between image x_i and the rest of the image in the batch is-

$$O(x_i)_b = \sum_{j=1}^{n} c_b(x_i, x_j)$$

$$O(x_i) = [O(x_i)_1, O(x_i)_2, \ldots\ldots\ldots., O(x_i)_B]$$

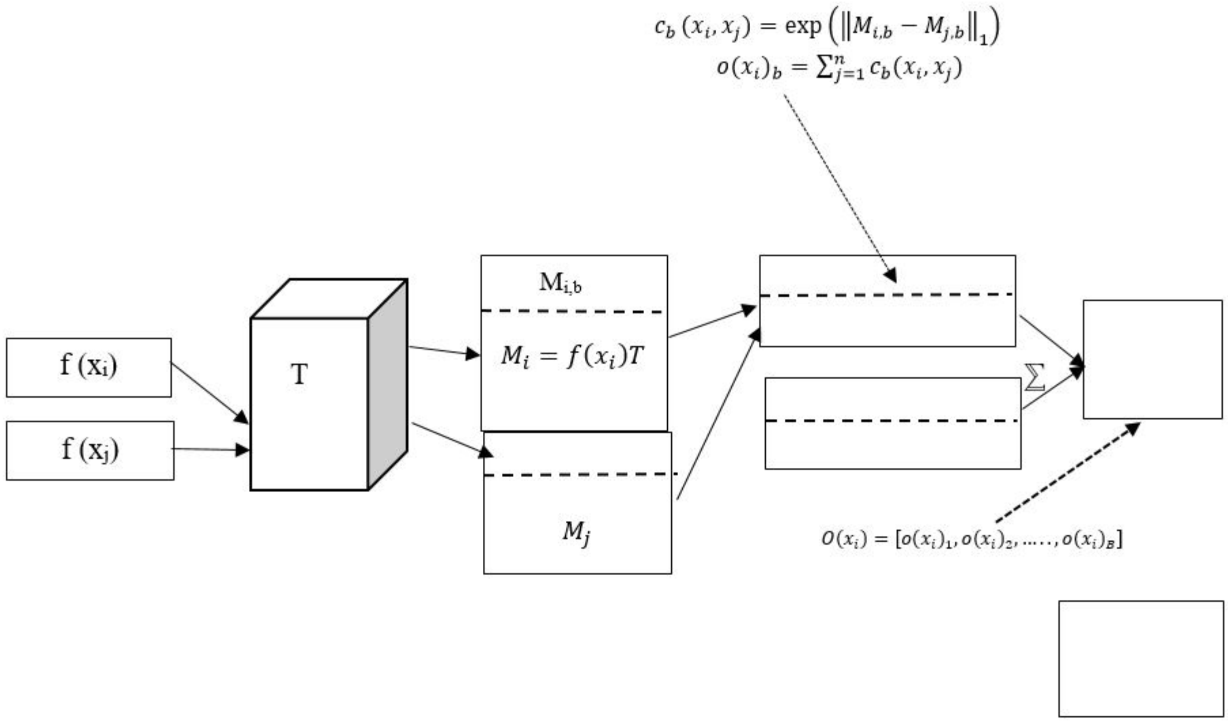


Figure 2.4 Full framework for Mini-batch discrimination

After computing the overall similarity, as per Figure 2.5, the discriminator can be guided to penalize the generator in the case of mode collapsing along with the task of detecting generated images.This similarity O(x) can be appended in one of the dense layers in the discriminator to classify whether the image is real or generated.

$$f(x_i), O(x_i)$$

In terms of generating visually appealing samples, this method can be regarded as superior to feature matching.

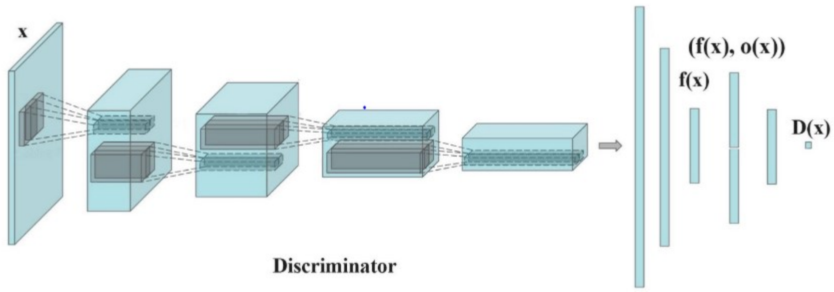


Figure 2.5 Discriminator framework in Mini-batch discrimination

2.7.3 Avoiding over-fitting by historical averaging

Historical averaging is a technique that keeps tracking of the parameters of the past. Then, it takes average of the parameters in the past and adds an L2 cost to the cost function to penalize model different from the historical average.

$$||\theta - \frac{1}{t}\sum_{i=1}^{t}\theta[i]||_2$$

According to [40], for GANs with non-convex objective function, historical averaging may stop models circling around the equilibrium point and act as a damping force to converge the model.

2.7.4 One-sided label smoothing

Label smoothing can reduce the risk of generating adversarial examples in GANs. Instead of classifying real and fake images through binary labels, we can smooth the target labels for both real and fake images. So, we can consider applying labels 0.9, 0.8, 0.7 and 0.1, 0.2, 0.3 to the images respectively.

2.7.5 Using labels (Conditional GAN)

In GAN, there is no control over the modes of the data to be generated. The conditional GAN changes that by adding label to the generator, thereby guiding the GAN training.

Labels can also be added to the discriminator input to distinguish real images better. Figure 2.6 depicts the framework for Conditional GAN.

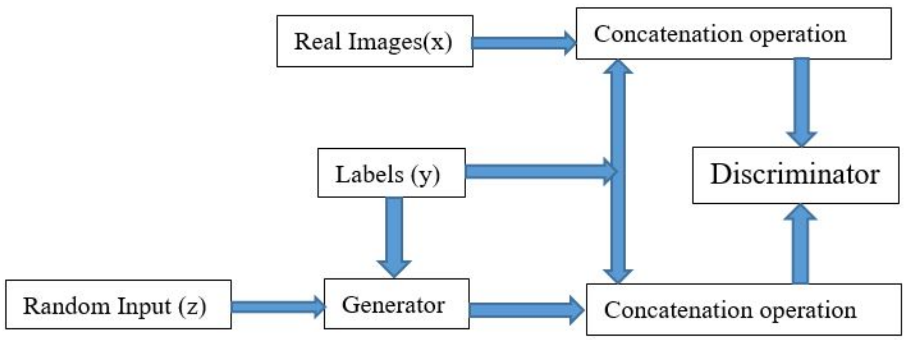


Figure 2.6 Framework for Conditional GAN

2.7.6 Cost functions

Cost function is one major research area in GAN, so research has been ongoing to find the best suitable cost functions on all the datasets. Considering that, Table 2.1 lists some of the cost functions for some common GAN models. In the table, for the sake of simplicity, we omit the use of any subscript in the expectation operator (E) in order to specify its data region, which we have previously used. Also, instead of using minmax operator in the GAN equation, in contrast to our previously described format, we separately describe the Generator and Discriminator equation in Table 2.1.

So, after having an insight about all these cost functions, we can infer that there is no single cost function that performs the best among all different datasets. It can also be concluded after comparing the FID score (lower the value, better the model) of different GAN models on some of the datasets as per Table 2.2 according to [41]. Along with these, good hyper-parameters and learning rate selection for both the generator and discriminator

Name	**Value Function**
GAN	$L_D^{GAN} = E[log(D(x))] + E[log((1 - D(G(z)))]$ $L_G^{GAN} = E[log(D(G(z))]$
LSGAN	$L_D^{LSGAN} = E[(D(x) - 1)^2] + E[D(G(z))^2]$ $L_G^{LSGAN} = E[(D(G(z)) - 1)^2]$
WGAN	$L_D^{WGAN} = E[D(x)] - E[D(G(z))]$ $L_G^{WGAN} = E[D(G(z))]$ $W_D \leftarrow \text{clip by value}(W_D, -0.01, 0.01)$
WGAN-GP	$L_D^{WGAN-GP} = L_D^{WGAN} + \lambda E[(\lvert\nabla D(ax - (1 - \alpha G(z)))\rvert - 1)^2]$ $L_G^{WGAN-GP} = L_G^{WGAN}$
DRAGAN	$L_D^{DRAGAN} = L_D^{GAN} + \lambda E[(\lvert\nabla D(ax - (1 - \alpha x_p))\rvert - 1)^2]$ $L_G^{DRAGAN} = L_G^{GAN}$
CGAN	$L_D^{CGAN} = E[log(D(x, c))] + E[log((1 - D(G(z), c))]$ $L_G^{CGAN} = E[log(D(G(z), c)]$
infoGAN	$L_{D,Q}^{infoGAN} = L_D^{GAN} - \lambda L_1(c, c\prime)$ $L_G^{infoGAN} = L_G^{GAN} - \lambda L_1(c, c\prime)$
ACGAN	$L_{D,Q}^{ACGAN} = L_D^{GAN} + E[P(class = c\backslash x)] + E[P(class = c\backslash G(z))]$ $L_G^{ACGAN} = L_G^{GAN} + E[P(class = c\backslash G(z))]$
EBGAN	$L_D^{EBGAN} = D_{AE}(x) + max(0, m - D_{AE}(G(z)))$ $L_G^{EBGAN} = D_{AE}(G(z)) + \lambda * PT$
BEGAN	$L_D^{BEGAN} = D_{AE}(x) - k_t * D_{AE}(G(z))$ $L_G^{BEGAN} = D_{AE}(G(z))$ $k_{t+1} = k_t + \lambda(\gamma * D_{AE}(x) - D_{AE}(G(z)))$

Table 2.1 Cost functions assoiated with different types of GANs

are required. A carefully tuned GAN models may mitigate some serious GAN's problems like mode collapse, stability issue.

Variants	**MNIST**	**FASHION**	**CIFAR**	**CELEBA**
MMGAN	9.8 ± 0.9	29.6 ± 1.6	72.7 ± 3.6	65.6 ± 4.2
NSGAN	6.8 ± 0.5	26.5 ± 1.6	58.5 ± 1.9	55.0 ± 3.3
LSGAN	7.8 ± 0.6	30.7 ± 2.2	87.1 ± 47.5	53.9 ± 2.8
WGAN	6.7 ± 0.4	21.5 ± 1.6	55.2 ± 2.3	41.3 ± 2.0
WGAN-GP	20.3 ± 0.5	24.5 ± 2.1	55.8 ± 0.9	30.0 ± 1.0
DRAGAN	7.8 ± 0.4	27.7 ± 1.2	69.8 ± 2.0	42.3 ± 3.0
BEGAN	13.1 ± 1.0	22.9 ± 0.9	71.4 ± 1.6	38.9 ± 0.9
VAE	23.8 ± 0.6	58.9 ± 1.2	155.7 ± 11.6	85.7 ± 3.8

Table 2.2 FID score associated with different types of GANs for different datasets from [41]

In final, it can be concluded that GAN models are becoming popular after solving issues associated with the training process. We know that deep learning models do not introduce much diversity in terms of the design of the architecture. So,after solving convergence and stability concerns of GAN, shifting the research toward introducing diversity in the design of the GAN architecture will be much more challenging task and insightful too.

2.8 Insight about CGAN

As our work is associated with the architecture of Conditional Generative Adversarial Network (CGAN), we would like to delve into its architectural framework, thereby paving the way to guide to our work. It is a supervised type of GAN that involves the conditional generation of samples by a generator. GAN can be extended to a conditional model if both the generator and discriminator are conditioned on some extra information. In essence, the conditioning can be done by feeding $\boldsymbol{y}$ into both the discriminator and generator as an extra part in loss function. Basically, in an unconditioned generative model, there is no control on modes of the data being generated. However, by conditioning the model on additional information it is possible to direct the data generation process. Such conditioning could be based on class labels, on some part of data for inpainting, or even on data from different modality [42]. To be very broad, Image-to-image translation problems can often be formulated as per-pixel classification or regression. These formulations treat the output space as unstructured in the sense that each output pixel is considered conditionally independent from all others given the output image. Conditional GANs instead learn a structured loss. Structured losses penalize the joint configuration of the output. Inspired from these kind of behavior of CGAN model, popular image to image translation network named Basic Pix2Pix framework has been developed in such a way that loss is learned here and any possible structural difference between target and generated output can be penalized. Its two most inevitable part, generator and discriminator architectures differ from the traditional design of generator and discriminator in the GAN structure. In this framework, generator has been designed via U-Net based architecture as we will describe in the subsequent chapter, whereas convolutional PatchGAN classifier [43] is used in the discriminator design.

2.8.1 Basic Architecture

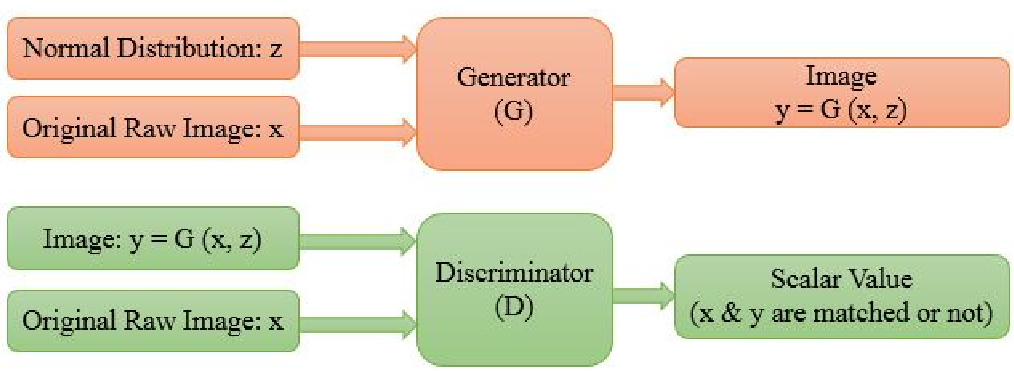


Figure 2.7 Different parts of conditional generative adversarial network

For the deep insight about this architecture, the basic architecture has been shown in

Figure 2.7. Here, without z, the network could still learn a mapping from x to y, but only tends to produce deterministic outputs thereby failing to match any distribution other than a delta function. Instead, in its final implementation, noise has only been introduced in the form of dropout such that the transition from the domain of original raw image to the domain of translated image will be much smoother.

2.8.2 Methodology

In contrast to GANs which learn a mapping from random noise vector z to output image $y : G : z \rightarrow y$, CGANs learn a mapping from observed image x and random noise vector z, to $y : G : x, z \rightarrow y$. The objective of a conditional GAN can be expressed as-

$$L_{CGAN}(G, D) = E_{x,y \sim p_{data}(x,y)}[logD(x, y)] + E_{x \sim p_{data}(x), z \sim p_z(z)}[log(1 - D(x, G(x, z))]$$

Where G tries to minimize this objective against an adversarial D that tries to maximize it, i.e. $G^* = \arg min_G \max_D L_{CGAN}(G, D)$. From the previous experience [44], it is apparent that GAN objective should be mixed with L2 distance loss as a regularizer to converge the two conflicting model. Also, due to the problem of having blurring effect after using L2 distance loss in the final translated image, it is beneficial to use L1 distance loss.

$$L_{L1}(G) = E_{x,y \sim p_{data}(x,y), z \sim p_z(z)}[||y - G(x, z)||]_1$$

So our final objective function should be:

$$G^* = \arg min_G \max_D L_{CGAN}(G, D) + \lambda * L_{L1}(G)$$

As evident from [7], this Pix2Pix framework has been successfully experimented on many computer vision tasks, such as Semantic Labels $\leftrightarrow$ photo trained on Cityscapes dataset, Architectural labels $\rightarrow$ photo trained on the CMP facades dataset, Map $\leftrightarrow$ aerial photo trained on data scraped from Google Maps, BW $\rightarrow$ color photos, Edges $\rightarrow$ photo, Day $\rightarrow$ Night etc. Also, the success of this model is to produce superior results even in the case of small training dataset. The following figures present some illustration of this Pix2Pix framework on natural images. Based on this framework, we will design our proposed

Figure 2.8 Example results on Google Maps at 512x512 resolution

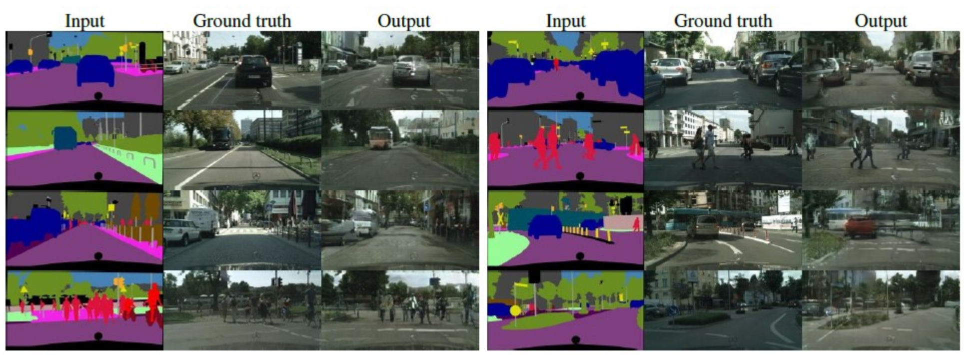


Figure 2.9 Example results on Cityscape Labels → photo, compared to ground truth

framework in the chapter 4.

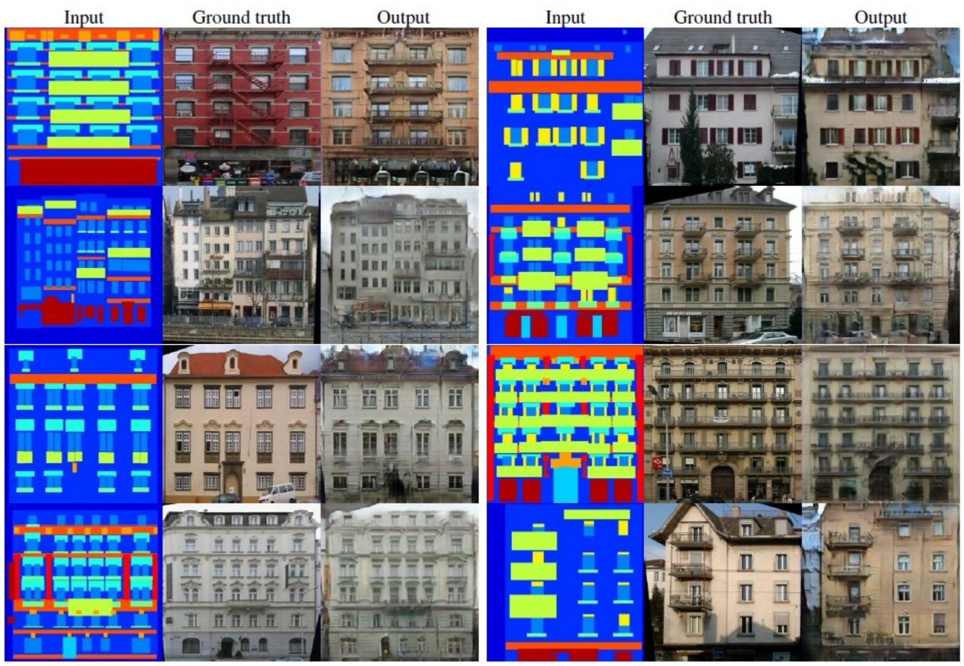


Figure 2.10 Example results of our method on facades labels → photo, compared to ground truth be implemented via Pix2Pix framework along with our proposed upgradation methodology to Pix2Pix framework and also it will be shown how both the methods surpass state of the art result in the field of medical imaging segmentation task.

Chapter Three

Biomedical Image Segmentation via Deep Learning

3.1 Evolution of Deep Learning in Medical Imaging

Considering the success of deep learning approaches into many computer vision tasks, it seems quite imminent to introduce the deep learning methods of computer vision in medical imaging applications. Initially from the 1970s to 1990s, after the advent of scanning and loading medical images into computer, automated medical image analysis included the sequential application of low level pixel processing (edge and line detector filters, region growing) and mathematical modeling (fitting lines, circles and ellipses) to construct compound rule based systems that solved particular tasks. At the end of 1990s, after training data was becoming popular to upgrade automated system then, supervised techniques such as active shape models for image segmentation [45], atlas methods (where the atlases that are fit to new data, form the training data), the concept of feature extraction and the use of statistical classifiers for computer aided detection and diagnosis. So this really indicates the shift from human designed systems to systems trained by computers using example data from which feature vectors are extracted.

The next step is to train computers in a way to learn features that optimally represent the data for the problem at hand. Much of the actual effort in deploying machine learning algorithms goes into the feature engineering which is too much labor intensive and also highlights the weakness of the existing machine learning algorithms. The most successful type of image modeling network to date is Convolutional Neural Networks (CNNs) were first applied to medical image analysis by 1995 [46]. But due to lack of computing power of the computers, both the real world and medical imaging application did not gather momentum through the use of CNNs. But its major contribution has been visible to world after 2012, when the proposed CNN, called AlexNet, won the ImageNet competition by a large margin [47]. However, the medical image analysis community has taken notice of these huge pivotal developments due to transition from handcrafted features to analyzing complex deep features. Because, many existing medical image processing methods rely on morphological feature representations to identify local anatomical characteristics. However, such feature representations were designed mostly by human experts. So, the image features, assume for 1.5-T T1-weighted brain MR images, are not applicable to work for other image types such as 7.0-T T1-weighted MR images. The superior performance of the deep learning based optimally extracted features can be evident from the comparison found in [48] through the process of absorbing the feature engineering step into the learning step. [49] includes some of these representative and automatically designed features such as layer-wise stacking of feature extraction, sparse coding scheme along with the inference of Maximum A Posteriori value etc. After the success of deep features in medical imaging applications, the number of papers grew rapidly after 2015. The development after advent of deep feature extraction process has been represented graphically in [50], which also includes a survey on 300 research papers regarding deep learning approaches in medical imaging applications.

3.2 Deep Learning in Medical Imaging Applications

Due to high resolution of medical images and unavailability of large annotated medical imaging datasets, traditional methods in medical imaging analysis suffer from poor model generalization ability. Deep learning based methods, the most breathtaking branch of the machine learning field, provide the most effective way to construct an end to end model in pursuit of Computer Aided Diagnosis (CAD) systems. However, the applications of deep models in the medical image analysis domain require great effort to catch up with other areas of imaging because deep architectures require large datasets to obtain outstanding features. However, medical images are usually difficult to acquire and thus, medical datasets are typically relatively small. Therefore, the approach is apt to lead to overfitting of the model if we directly use deep model to address a small dataset. Instead, deep learning based medical imaging analysis has been improvised in such a robust way that it overcomes the barrier of so called small datasets as well as overfitting issue. Generally, deep-learning pipelines for medical image analysis comprise many interconnected and common components. These include-

- Separation of data into training, testing and validation sets;
- Randomized sampling during training;
- Image data loading and sampling;
- Data augmentation;
- A network architecture defined as the composition of many simple functions
- A fast computational framework for optimization and inference;
- Metrics for evaluating performance during training and inference.

Recent reviews [48], [50] have highlighted that deep learning has been applied to a wide range of medical image analysis tasks such as deep feature representation learning in medical images, detection of structures (organs, body parts, different cell etc.) for segmentation, for Computer Aided Detection, for Computer Aided Diagnosis, for image registration, reconstruction etc. As this work is solely focused on the segmentation task regarding deep learning approaches, in the subsequent section, we will discuss broadly about the deep learning approaches in biomedical image segmentation task.

3.3 Medical Image Segmentation via deep learning

The principal modalities in medical image analysis are Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). The CT image can diagnose muscle and bone disorders, such as bone tumors and fractures, while the MR image can offer a good soft tissue contrast without radiation. Functional images, such as PET, lack anatomical characterization, which can provide quantitative metabolic and functional information about diseases.

Lung cancer is by definition a malignant tumor early detection of which could reduce the mortality rate and increase the patient's survival rate. CT imaging is an efficient medical screening test used for lung cancer diagnosis and detection. With the help of CAD sytem, the physicians analyze and diagnose the lung tissues. However, designing an effective lung segmetation method is a challenging problem, especially for abnormal lung parenchyma tissue, where the nodules and blood vessels need to be segmented with the lung parenchyma. Moreover, the lung parenchyma needs to be separated from the bronchus regions that are often confused with the lung tissue. Here, the proper utilization of popular image segmentation architecture, U-Net has eased the path to malignant lung tissue segmentation [51]. Brain extraction, specifically skull stripping, from magnetic resonance imaging (MR) data is an essential step in many neuroimaging applications, amongst other surgical planning, cortical structure analysis, surface reconstruction etc. It has been reported in [52] that anatomical variability, as well as age and the extent of brain atrophy, e.g. due to neurodegeneration have influenced the results of brain extraction methods. When considering the MR brain scans for the purpose of brain tumors, the problem becomes even more severe. 3D Deep Convolutional Neural Network (CNN) has been implemented in [53] in such way that requires minimal to no parameter tuning and still handle images from clinical routine i.e. from a wide age range, possibly including artifacts, contrast agent and pathologically altered brain tissue. Also, it facilitates the use of both single modality and the combination of several modalities (e.g. T1, T2, T2-FLAIR etc.). While training their 3D CNN, they constructed mini batches of multiple cubes that were larger than the actual input to their 3D CNN for computational efficiency. Again, it utilized the concept of FCNNs, so the output could be the block of predictions per input, rather than one single output. [54] presented a method using CNNs for the segmentation of three tissue types: white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF), in MR brain images of 6-8 months old infants, which have low contrast between WM and GM. Automatic segmentation of MR brain images into a number of tissue classes has been implemented in [55] through the use of multiple patch sizes and multiple convolution kernel sizes to acquire multi scale information about each voxel. This method allows to obtain accurate segmentation details as well as spatial consistency which also omits the use of explicit spatial features in contrast to the previous work [56] in the brain segmentation task.

The most well-known architecture in medical image analysis, specifically image segmentation task is U-net [9]. The two most architectural contribution of U-Net are the combination of an equal amount of upsampling & downsampling layers and the introduction of skip connections between the upsampling & downsampling layers. In the upsampling layers, there are a large number of feature channels, which allow the network to propagate contextual information to higher resolution layers. As a consequence, the expansive path is more or less symmetric to the contracting path yielding a u-shaped architecture. So it facilitates the processing of full images in one forward pass in contrast to patch based CNNs [57]. Figure 3.1 represents the trademark architecture of U-Net. Also, this work can be extended to 3D image segmentation task [58] by feeding U-Net with a few 2D annotated slice from the same volume. Moreover, there have been 3D variant defined over U-Net called V-Net [59] by redefining the objective function based on Dice coefficient.

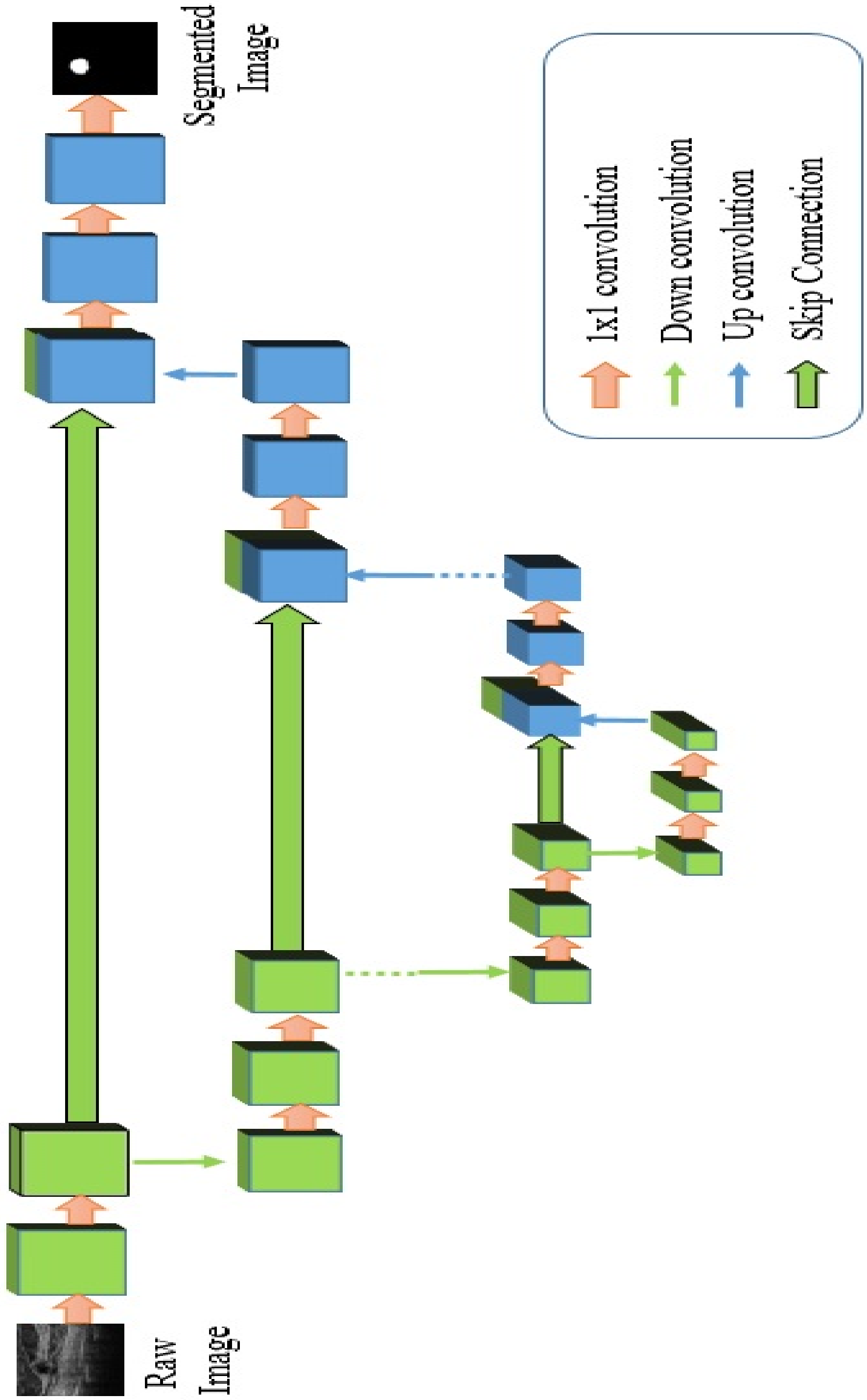


Figure 3.1 U-net architecture

3.3.1 Generative Adversarial Network in Medical Imaging Analysis

Recently, Generative Adversarial Network (GAN) has been emerged as a very strong and popular tool for medical imaging synthesis due to the scope of data augmentation to alleviate the data scarcity and overfitting problem in medical imaging applications and also the advantages regarding adversarial training schemes. The basic idea behind this network has already been discussed in the previous chapter. It has already emerged as an auto choice in computer vision tasks. After analyzing from internet, only one review paper [60] regarding GAN in Medical Imaging Analysis has been found to date. According to [60], GANs can be used in medical imaging applications in two ways; the first one is focused on the generative aspect, which can help in exploring and discovering the underlying structure of training data and learning to generate new images. The second one focuses on the discriminative aspect, where the discriminator can be regarded as a learned prior for normal images so that it can be used as a regularizer or a detector when presented with abnormal images.

As we have previously discussed, U-Net architecture has provided the state of the art results in biomedical image segmentation tasks. Still it lacks the guarantee of spatial consistency in the final segmentation map. Traditionally, conditional random field and graph cut methods are usually adopted for segmentation refinement by incorporating spatial correlation. Their limitation is that they only take into account pair wise potentials which might cause serious boundary leakages in low contrast regions, thereby reducing the sharpness in the final segmentation map. Another recently introduced paper in [61] implements U-Net after some modification to the original U-Net architecture, like the case of addition of dropout layers, binary cross entropy function being expanded to ternary cross entropy function. But still the representation does not seem very much convincing, as it does not ensure the model's generalization capability. In contrast to that, problem of refinement in the final segmentation map has been alleviated in [62] by introducing adversarial loss along with the encoder decoder based architecture. It helped to count on the fact of higher order potentials which formed the basis of considering the discriminator as a shape regulator. Laterly, this contribution has been extended to achieve the domain invariance [63], [64] such as between different modalities, image protocols etc. In work [65], extracted features coming from different depths has been compared in the discriminator through multi scale L1 loss. It imposed multi scale spatial constraints on the final segmentation map and the system achieved state of the art performance in the BRATS 13 and 15 challenges. In work [66], segmentation task has been performed in semi supervised way. In the annotated images, both the element wise loss and adversarial loss has been applied, whereas the unannotated images are only used to compute final segmentation map in a way to confuse the discriminator. Also, the network invariance to small perturbations of the training samples in order to reduce overfitting has been implemented through the adversarial training scheme [67]. Recently, a popular architecture named SegNet-cGAN has been successfully implemented in both for mammograms as well as X-ray based breast lesion segmentation task [68] [69]. This architecture has its association with the popular supervised ways of GAN training named CGAN, which can be rooted to the basic of [7]. The basic architecture for SegNet-cGAN has been shown in figure 3.2.

3.4 Cases for Ultrasound Images

As our work is concerned with Segmentation of Breast Lesions from Ultrasound Images using Generative Adversarial Networks, before delving deep into our proposed method, it would be better to through some insights about Ultrasound images. In terms of imaging modalities, Ultrasound (US) has become one of the most recognized and powerful screening as well as diagnostic tool for physicians and radiologists. Over the decades, it has been widely demonstrated that US has several advantages over other imaging modalities such as X-ray, Magnetic Resonance Imaging (MRI) and computed tomography, including its non-ionizing radiation, portability, non-invasive nature, accessibility as well as cost effectiveness. In cur-

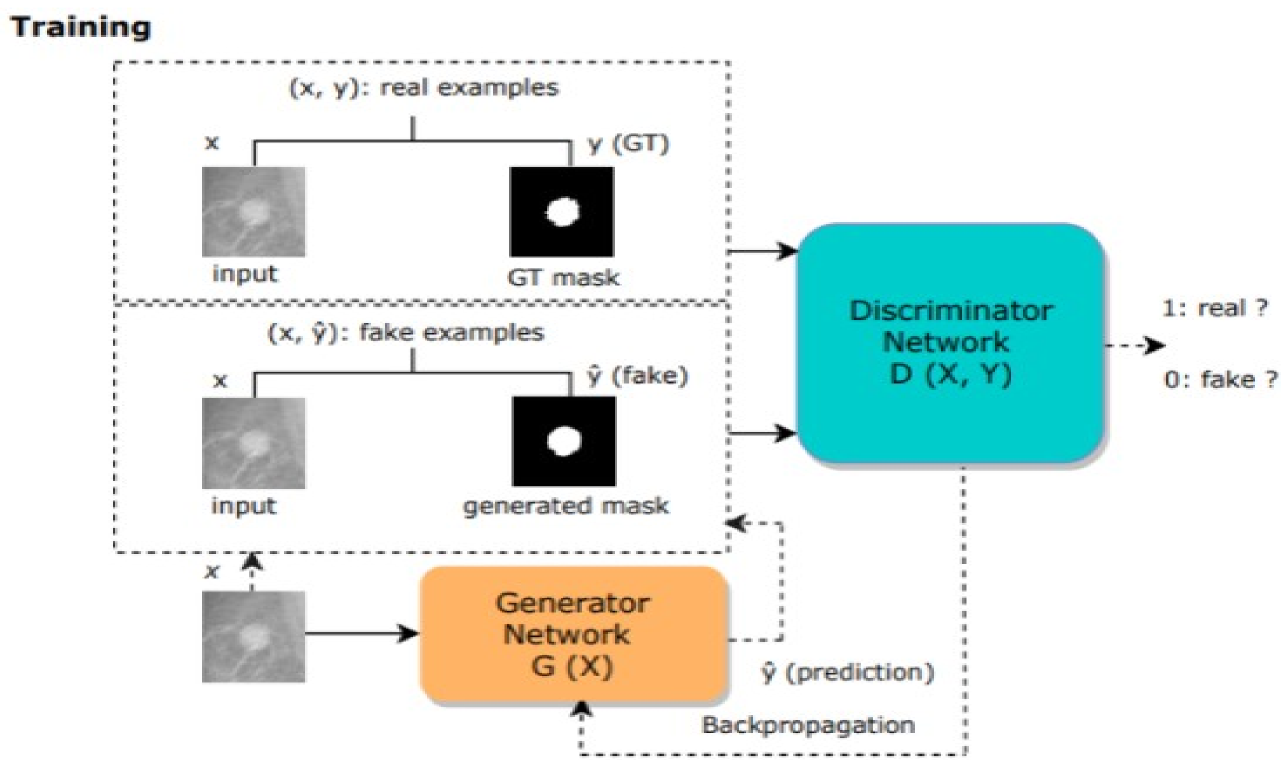


Figure 3.2 SegNet-cGAN Architecture

rent clinical practice, medical US has been applied to specialties such as echocardiography, breast US, abdominal US, transrectal US, intravascular US and prenatal diagnosis US, which is specially used in obstetrics and gynecology (OB-GYN). Whatever, US also imposes unique challenges, such as low imaging quality caused by noise and artifacts, high dependence on abundant operator experience, and high inter- and intra- observer variability across different institutes and manufacturer's US systems. For example, a study on the prenatal detection of malformations using US images demonstrated that the sensitivity ranged from 27.5% to 96% among different medical institutes [70]. So, in order to overcome these challenges, it is essential to develop advanced automatic US image analysis methods to make US diagnosis and assessment, as well as image guided interventions/ therapy, more objective, accurate and intelligent. After emerging as the top-most breakthrough technologies, deep learning has been adopted in developing Computer Aided Diagnosis system. A recently published review

[71] regarding deep learning in medical ultrasound analysis has mentioned many of the areas where deep learning has been adopted in a way to overcome the challenges in this image modalities and also discusses about the future challenges as well as room of improvement. Figure 3.3 illustrates the different stages of medical ultrasound analysis.

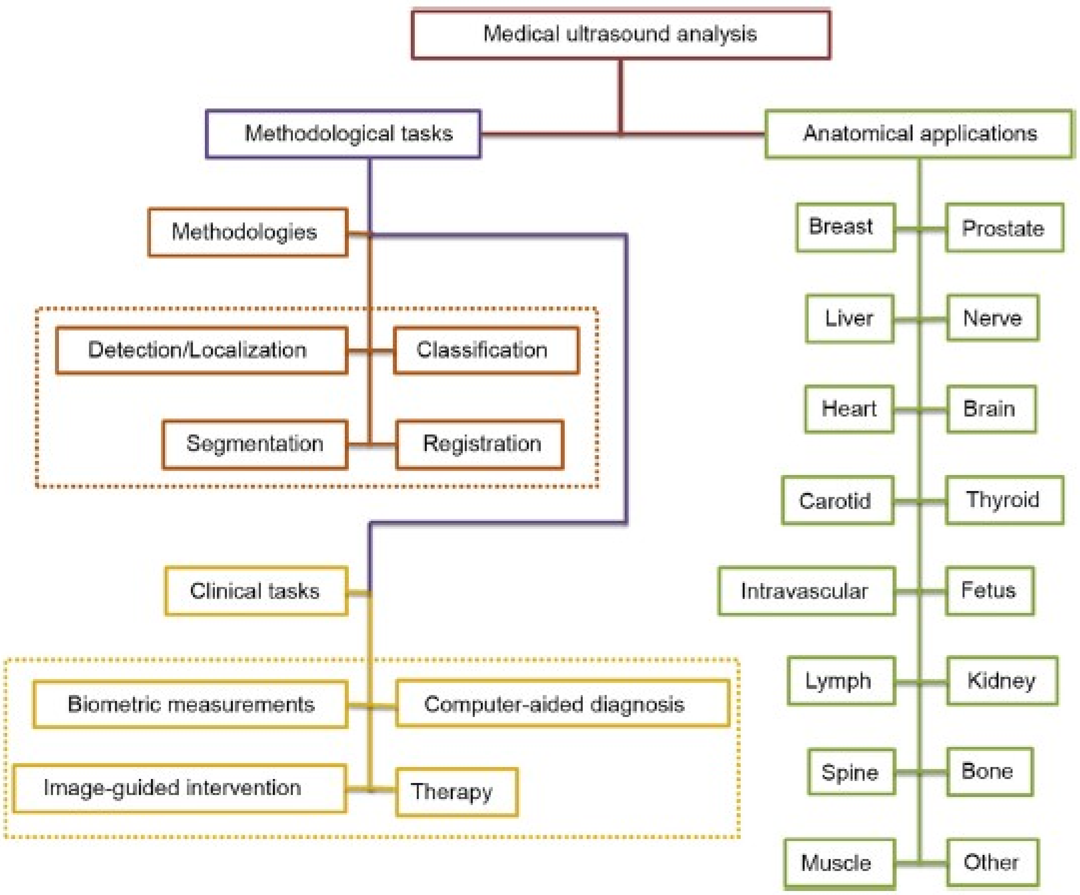


Figure 3.3 Illustration of Medical US Analysis

Chapter Four

Proposed Segmentation Method: Modified-CGAN (Mod-CGAN)

4.1 Multi-Task Learning

Multi-Task Learning (MTL) is a very effective area for deep learning applications, where limited training samples are trained for deep models. In MTL, there can be many different as well as general learning tasks such as supervised tasks (classification and regression), unsupervised tasks (clustering problems), semi supervised tasks, reinforcement learning tasks, multi view learning tasks etc. Among these learning tasks, all of them or atleast a subset of them are related to each other. After getting inspired from the fact that knowledge contained in a task can be leveraged by the other tasks, it can be inferred that learning these tasks jointly can lead to much better performance improvement compared to learning these tasks individually, thereby giving rise the scope of MTL. So, MTL can be mathematically characterized such that given m learning tasks $\{T_i\}_{i=1}^{m}$ where all the tasks or subset of them are related, MTL aims to improve the learning of a model for T_i by using the knowledge contained in all or some of the m tasks. In MTL, the relatedness can be defined in three ways: when to share, what to share and how to share.The 'when to share' issue is to make choices between single task and multi task models for a multi task problem. 'What to share' refers to the form consisting either of feature, instance and parameter through which knowledge sharing among all the tasks will occur. After determining 'what to share', 'how to share' specifies concrete ways to share knowledge among tasks. The majority of MTL studies mainly focus on feature based and parameter based methods. In connection with our work, we will take into consideration some details of feature based MTL studies.

In feature based MTL, the primary approach is- feature learning approach. The feature learning approach focuses on learning common feature representations for multiple tasks based on shallow or deep models, where the learned common feature representation can be a subset or a transformation of the original feature representation. Based on the relationship between the original feature representation and the learned one, one can further categorize into feature transformation and feature selection approach respectively. The feature transformation approach is the learned features different from the original feature representations as these features are linear or non-linear transformations of the original feature representations. On the other hand, in feature selection approach, the learned feature representations are similar to the original features by eliminating useless features based on different criteria.

In many real life applications, some of the input matrices or output vectors can be shared among different tasks. Based on whether the inputs or outputs are shared among different tasks, MTL problems can be categorized into three different models, namely Multi-input single-output (MISO), Single-input multi-output (SIMO) and Multi-input multi-output (MIMO). All these three categories have different applications based on how it relates the input data to the output target. For the sake of analyzing the correlation of MTL with our task, we will only be concerned about SIMO which paves the way to interpret our work quite satisfactorily.

In the case of SIMO, single data source will be utilized for different tasks. There will exist a shared/ common network for all the tasks in hand. Each task can be defined in such a way that every task will try to benefit from the shared / common network. The case arises when we intend to have the segmentation, classification, detection all available in the final inference problem. The graphical representation of SIMO has been shown in Figure 4.1. The mean square loss formulation for the data fidelity term in this case is given as:

$$L(x, Y, W) = \sum_{s}^{S} ||x * w^s - y^s||_F^2$$

where $Y = \{y^1, y^2, ..., y^S\}$ denotes the set of multi-output inference, x denotes the single data source, and $W = [w^1 w^2 ... w^S]$ denotes the weight matrix with its s-th column w_s denoting the weight vector corresponding to the mapping of x to y^s.

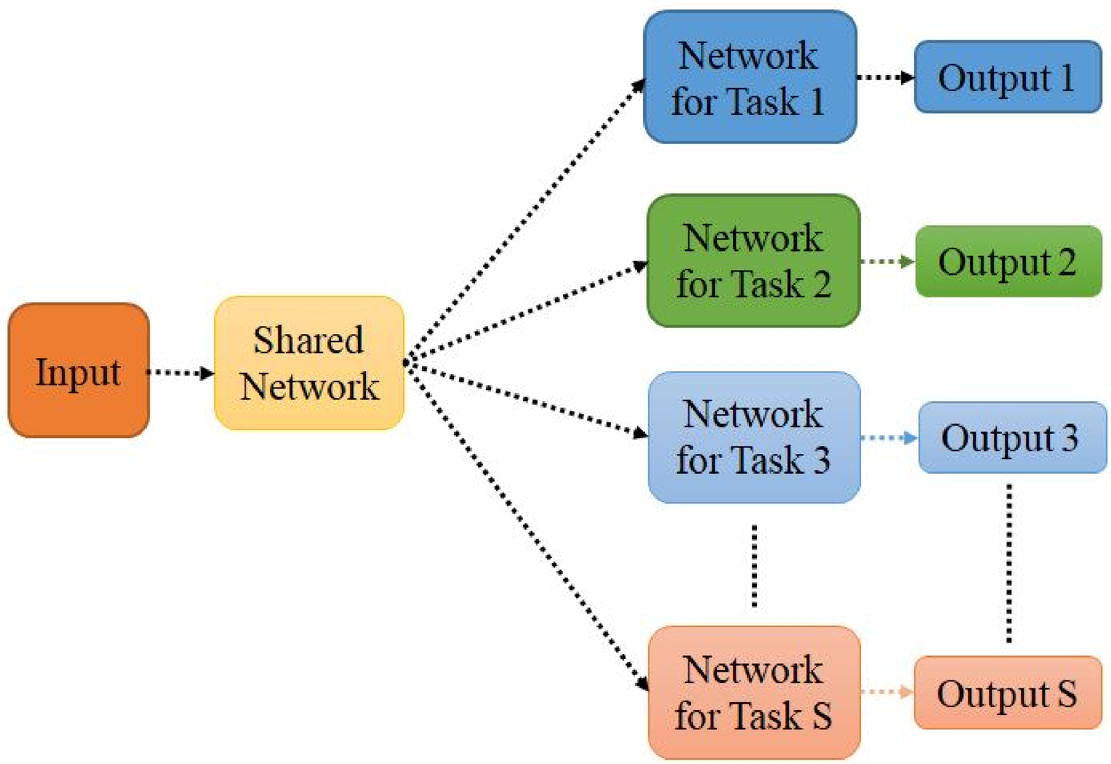


Figure 4.1 single-input multi-output (SIMO), where single set of input are mapped to multiple target

4.2 Proposed Method: Modified - CGAN (Mod-CGAN)

As depicted from chapter 2, Pix2Pix framework has been designed on the theory of conditional GAN. It is one kind of supervised learning when making use of adversary during learning. Our goal here is to generate the image in a domain which is different from the input domain. As we have our translated/ annotated images available, so we can utilize these images more by taking a sample layer output from the encoder part of the generator and compare it with the annotated images after taking it to the sample layer output domain to ensure the fact that generator is tracking its hunt for the right direction. This additional network has been termed as the Modifier network. This modifier network is actually the one which facilitates the use of Multi Tasking learning (MTL) in the CGAN framework. As the incorporation of MTL introduces a modification into the existing CGAN framework, the newly introduced network has been termed as Modified- CGAN, in short it can be termed as Mod-CGAN. Also, it is to be clarified that although this extra architecture has been operated in parallel with the generator during the training process only, it is not included in the test stage. Figure 4.2 represents the full architecture of this new proposed improvement to the Pix2Pix framework. In the modifier part, output represents the mask matching score which is ultimately a scalar value between 0 and 1. Mask matching score "1" interprets about the complete matching of features between raw image and ground truth image after those intermediate layers of operation. In the generator architecture, each encoder layer includes the functionality of down convolution, batch normalization and Relu activation whereas each decoder layer consists of transposed convolution, batch normalization, dropout, Relu activation and concatenation with corresponding encoder layer through skip connections. Here, the idea of skip connections has been undertaken from the architecture of previously described U-Net architecture. Also, it is to be clarified that the implementation of this new architecture do not incorporate the data augmentation technique here. The motivation here is that data augmentation technique includes stretching, shifting, rotating of the original images, which does not coincide with natural deformation of the body organs. Most of all, problem here is that it may bias the result in a random direction.

4.2.1 Basic Architecture

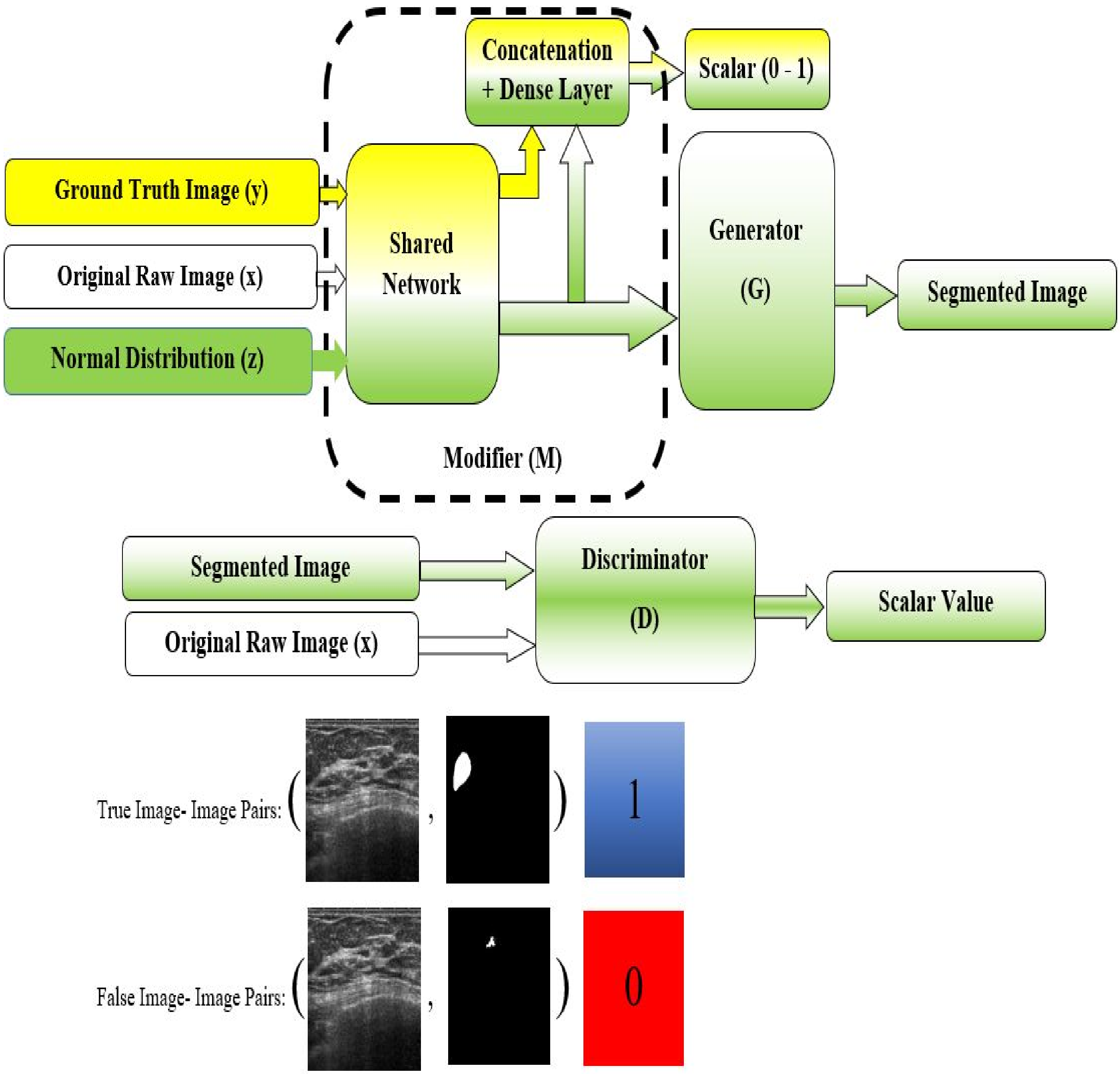


Figure 4.2 Basic architecture of Mod-CGAN

4.2.2 Training Structure

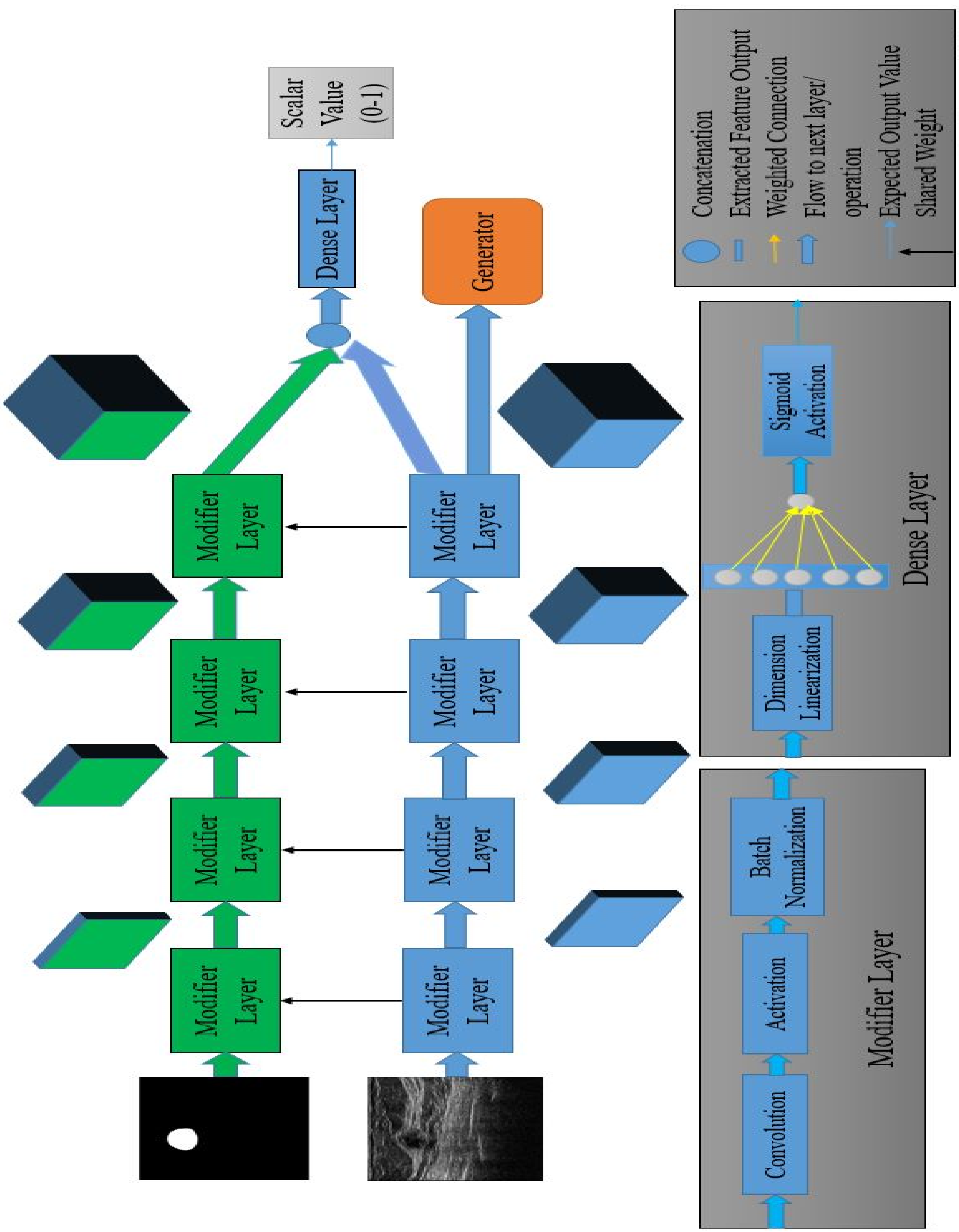


Figure 4.3 Detailed Modifier Architecture

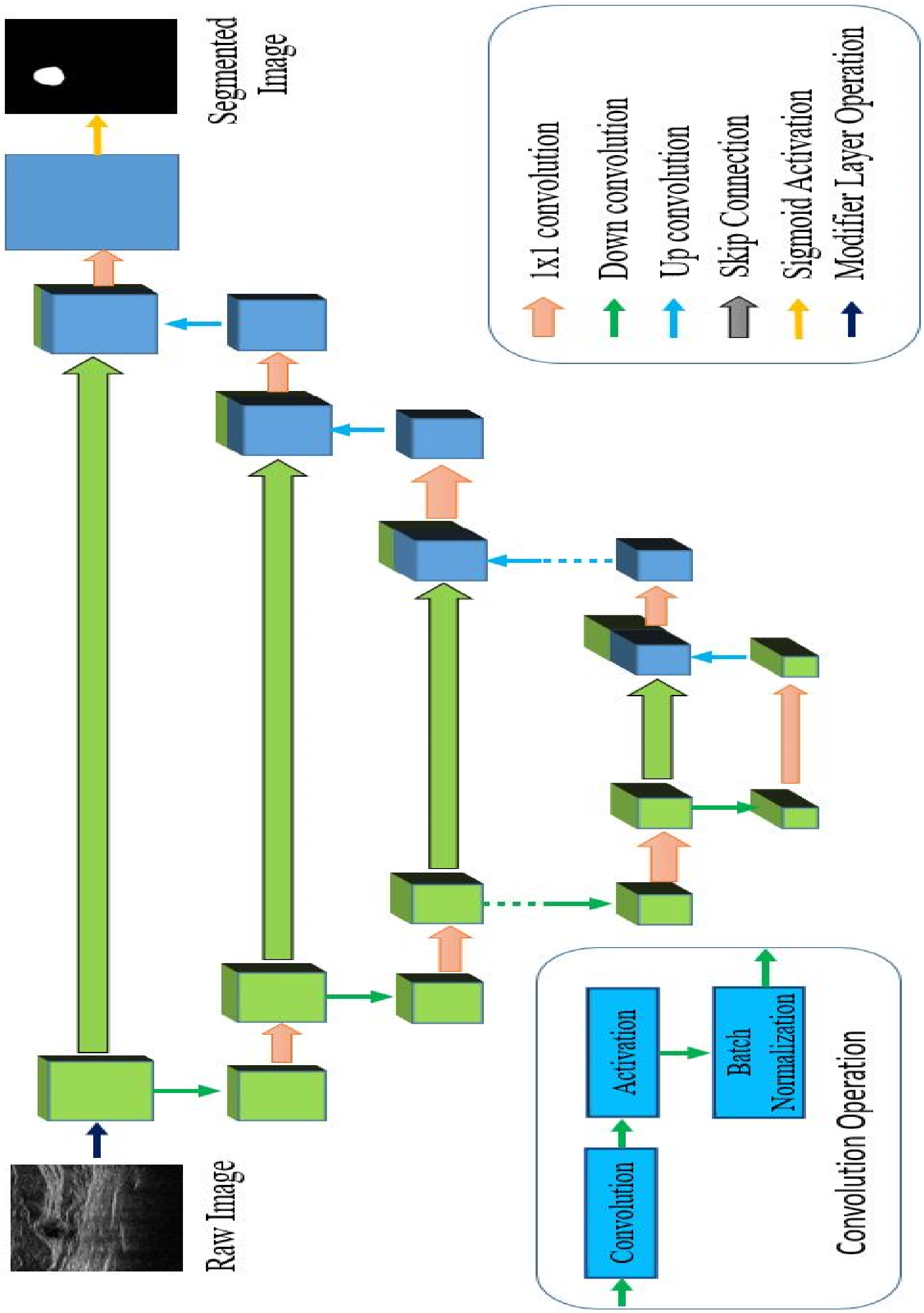


Figure 4.4 Detailed Generator Architecture

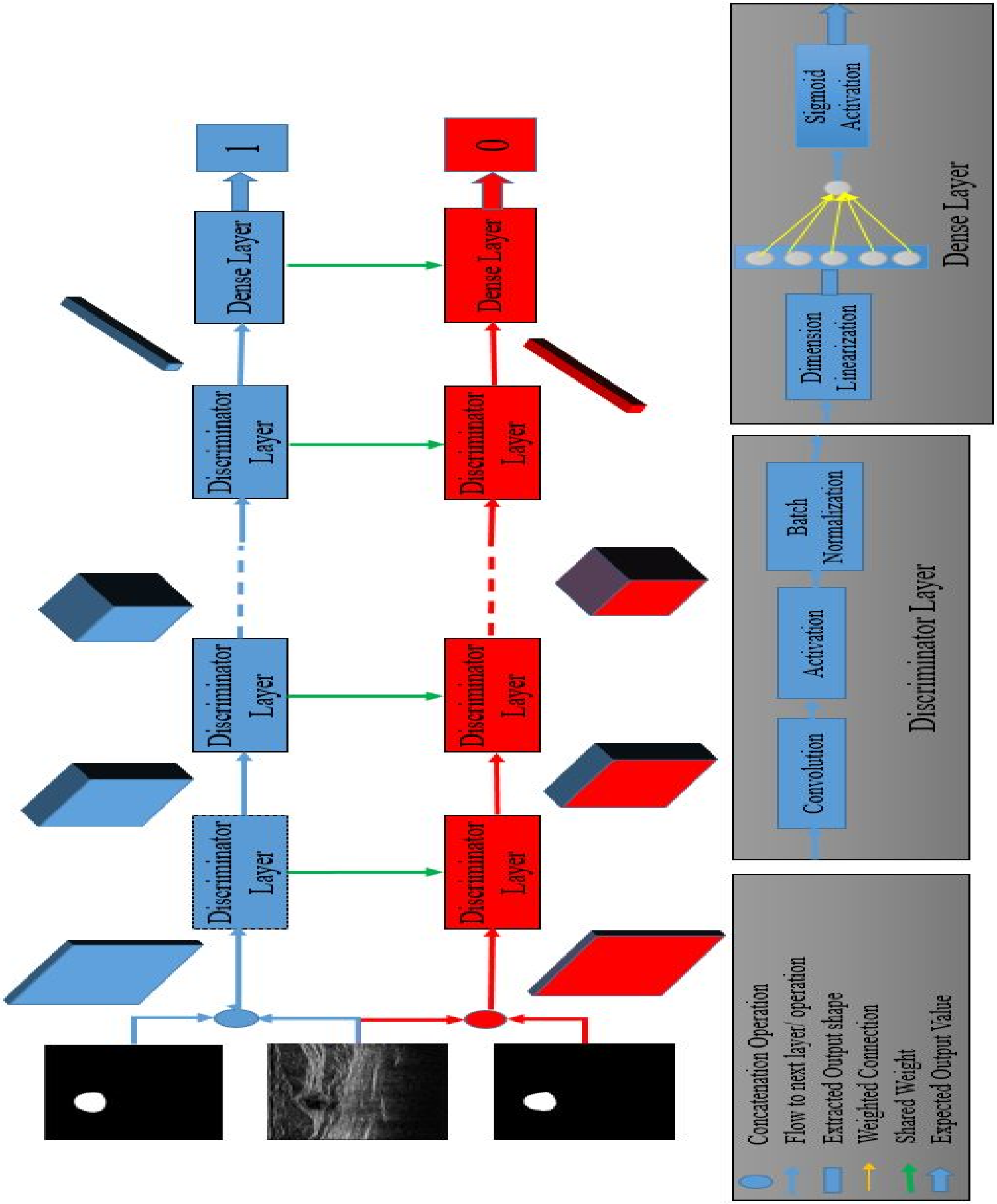


Figure 4.5 Detailed Discriminator Architecture

4.2.3 Test Structure

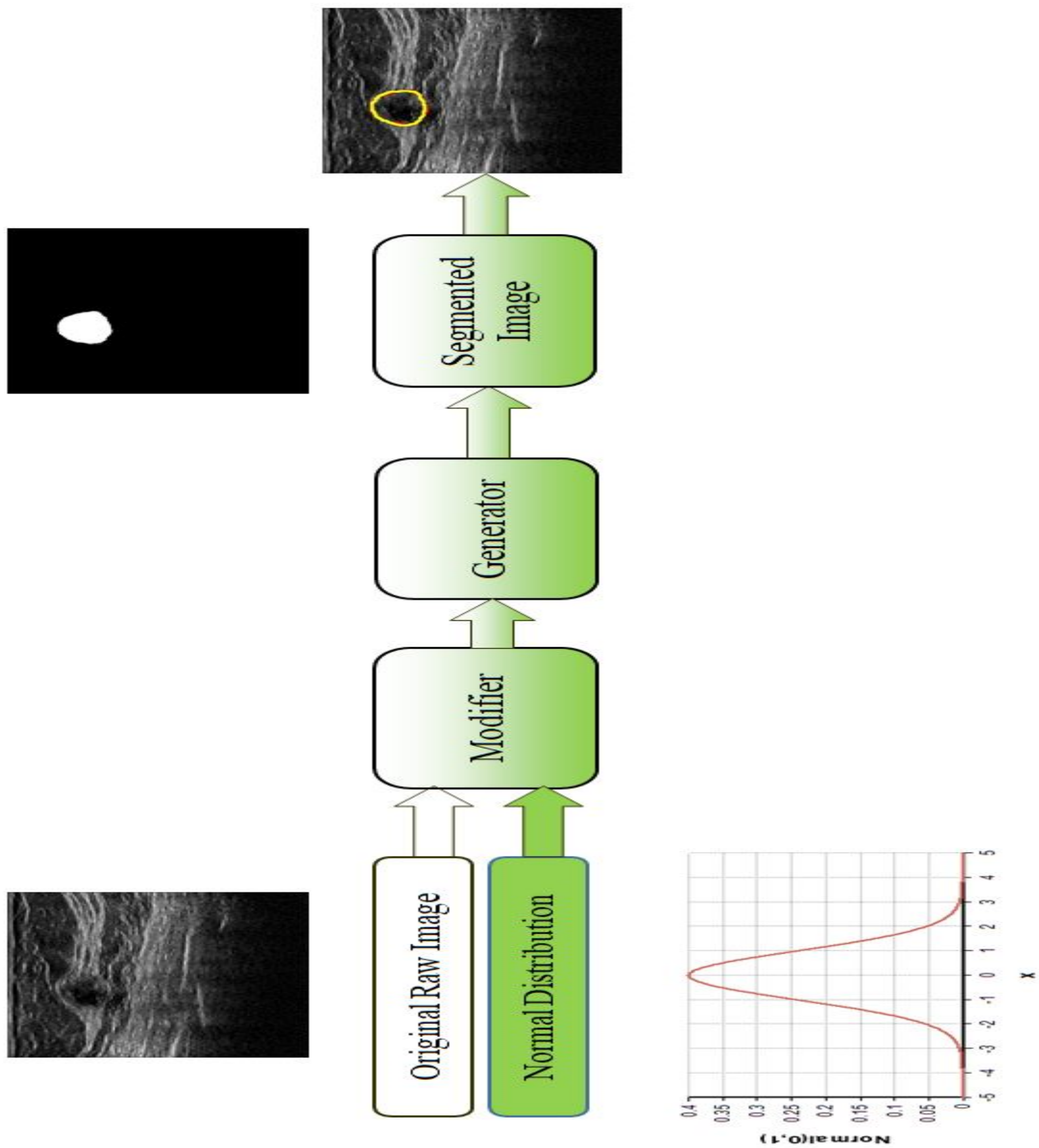


Figure 4.6 Architecture during test time

During the part regarding testing of the selected portion of our datasets, we have to utilize both the modifier and generator section, but in the modifier part, only the flow from original raw image has to be maintained due to not having the availability of ground truth annotated image. Segmented image will be the final generated image of our interest. Our model's performance will be assessed on the basis of final generated / segmented image. Figure 4.6 depicts the architecture during test time.

4.3 Features of Mod-CGAN

As evident from the previous section, Mod-CGAN consists of three networks, named as Modifier, Generator and Discriminator. Here, implementation for modifier network has been generalized so that actually it can be imposed on any encoder-decoder based architecture. Figure 4.3 depicts that the modifier network consists of some common initial layers of the encoder part from the generator. This modifier network is trained through the mask matching score. Generator and Discriminator network will remain same like the case of CGAN framework. Here, in order to implement the Mod-CGAN architecture, we do not require any extra additional information apart from raw image as well as ground truth image. Rather through this modifier network, more utilization of the available data can be ensured. Figure 4.7 and 4.8 clearly represent about the better visualization of Generator network between CGAN and Mod-CGAN architecture.

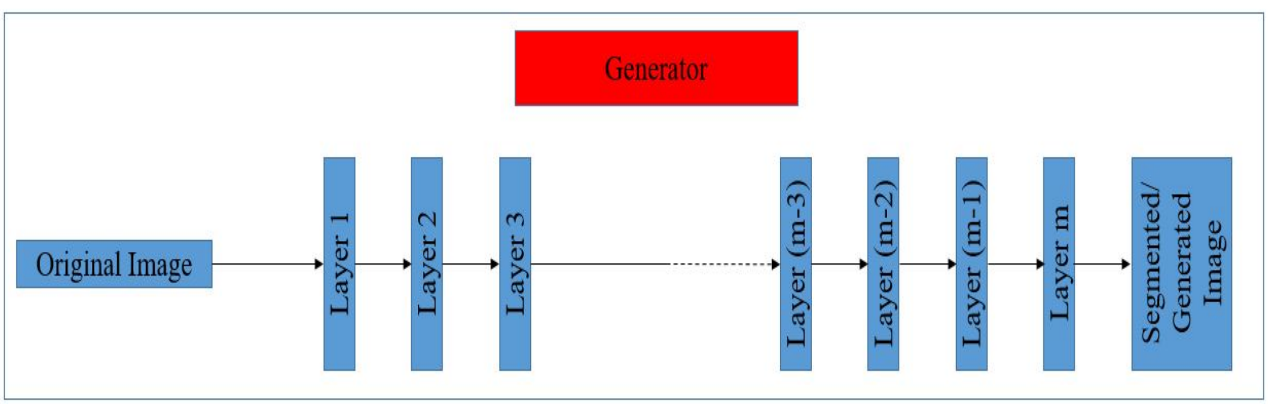


Figure 4.7 Diagram for Generator of CGAN

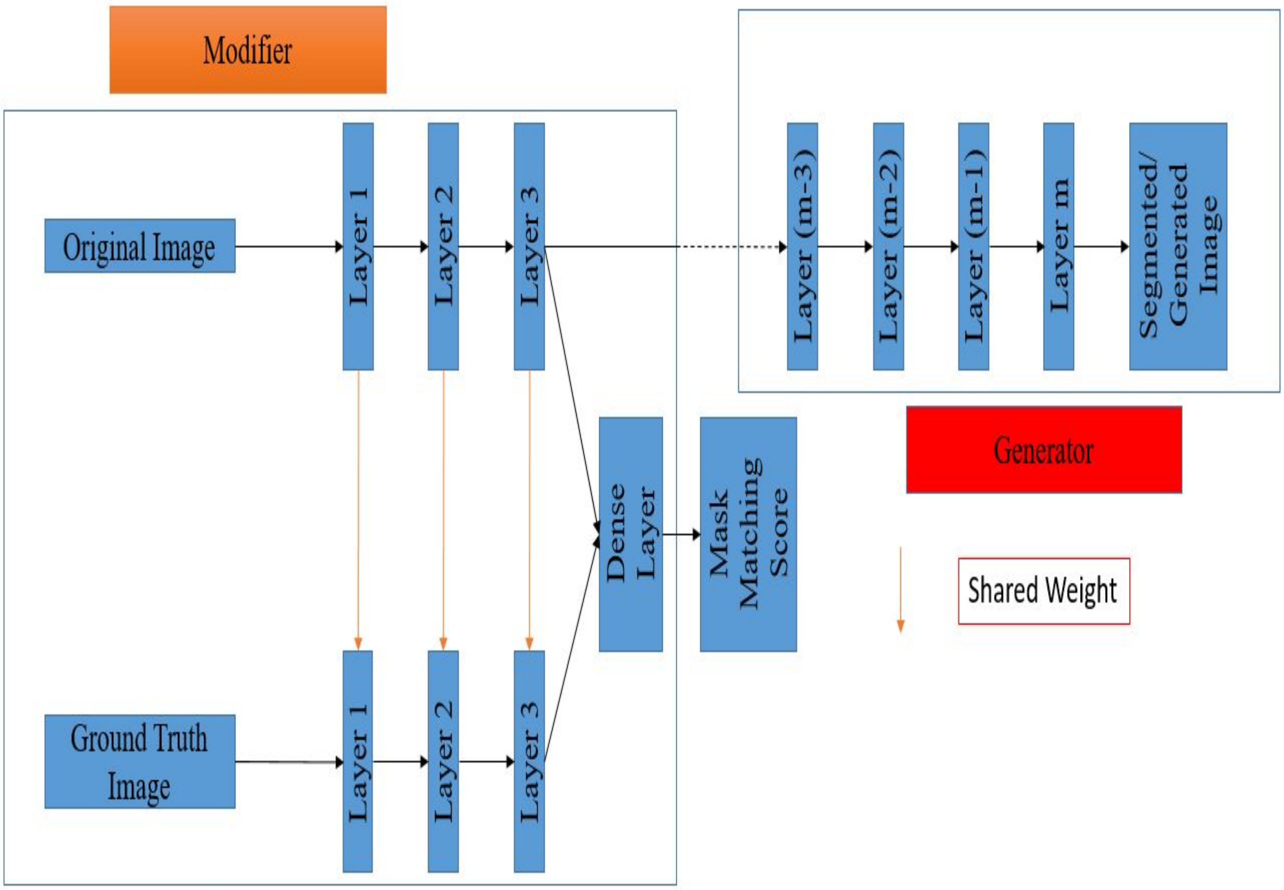


Figure 4.8 Diagram for Mod - CGAN

4.4 Objective Function

Here, our proposed architecture consists of three different architectures. We can define them as generator, discriminator and modifier architecture which is inclined within the generator architecture yet having different loss functions. The intention of generator is to fool the discriminator by its generated output, the discriminator trains itself from the pattern/ data structure of original ground truth output in order to deny the generator any chance to skip it. As the loss function of generator and discriminator are mutually exclusive, the issues of convergence still remain a big issue for GAN. In the case of conditional GAN, generator is trained in a supervised way. The modifier architecture has been developed in such a way that it guides the generator to the right direction. So, it is emerged from the part of the generator, yet having different loss function. For this case, the objective of a MTGAN can be written as -

$$L_{MTGAN}(G, D) = E_{x,y \sim p_{data}(x,y)}[logD(x, y)] + E_{x \sim p_{data}(x), z \sim p_z(z)}[log(1 - D(x, G(M(z, x/y)))]$$

Where G tries to minimize this objective against an adversarial D that tries to maximize it, i.e. $G^* = \arg min_G \max_D L_{MTGAN}(G, D)$. Also, as like the case in Pix2Pix framework, L1 distance loss has been utilized.

$$L_{L1}(G) = E_{x,y \sim p_{data}(x,y), z \sim p_z(z)}[||y - G(M(z, x/y))||]_1$$

So our final objective function for the part of MTGAN should be:

$$G^* = \arg \min_G \max_D L_{MTGAN}(G, D) + \lambda * L_{L1}(G)$$

Along with these conditional GAN part, also modifier network needs to be optimized in order to assist the generator to the path of convergence. Its final objective function can be defined simply as-

$$L_{modifier} = E_{x,y \sim p_{data}(x,y), z \sim p_z(z)}[logM(z, x/y)]$$

As in this part of the network, modifier has been designed conditionally such that available translated data have been more efficiently utilized. Modifier and GAN network will be trained independently, but during the execution of same training batch. In final, here, it can be said that it is a network of dual platform where adversaries has been introduced via the practice of conditional GAN network and idea of multi tasking learning have been utilized through the inclusion of modifier network. To the best of our knowledge, this is the first time, where biomedical image segmentation has been implemented through the exercise of dual network based architecture.

4.5 Distinction with Multi Tasking based GAN architecture

As the days go on, learning task for deep learning model will be more challenging such that there will be small room for the model to be generalized. Also, data scarcity always remains a concern for the model to upgrade to the next level. It refers to the fact that model's capacity is a key issue in terms of designing deep learning models for new researchers. Keeping it on mind, multi task based learning recently brings an extra addition to the existing deep learning models. As already discussed in Section 4.1, multi tasking based learning can rather be regarded as an art which can be followed through different way in different framework as required. That's why in literature, this multi tasking based learning has been incorporated in different formats despite being the same general name. In [72], multi task GANs has been utilized for performing two different tasks named Semantic Segmentation and Depth Completion Tasks with Cycle Consistency. The name multi task here is due to the fact that the outcome of Semantic Segmentation task has been fed as input to the Depth Completion task. Figure 4.9 depicts how the two tasks have been accomplished in the ideology of multi tasking. In [73], small object detection has been done via multi task Generative Adversarial Network. Here, multi task learning has been achieved completely in the discriminator network in the form of adversarial learning, classification learning and bounding box regression offset learning. Figure 4.10 depicts the situation. Figure 4.11 depicts the Mod-CGAN implementation on the basis of task.

From the figure shown in 4.9, 4.10, 4.11, it can be inferred that our proposed multi tasking based implementation is completely different from those available in the literature. Our proposed multi tasking based implementation works on several initial layers of the generator. For our case, only defined explicit task is lesion segmentation. That's why we have defined a new implicit but performance boosting task in the form of mask matching score.

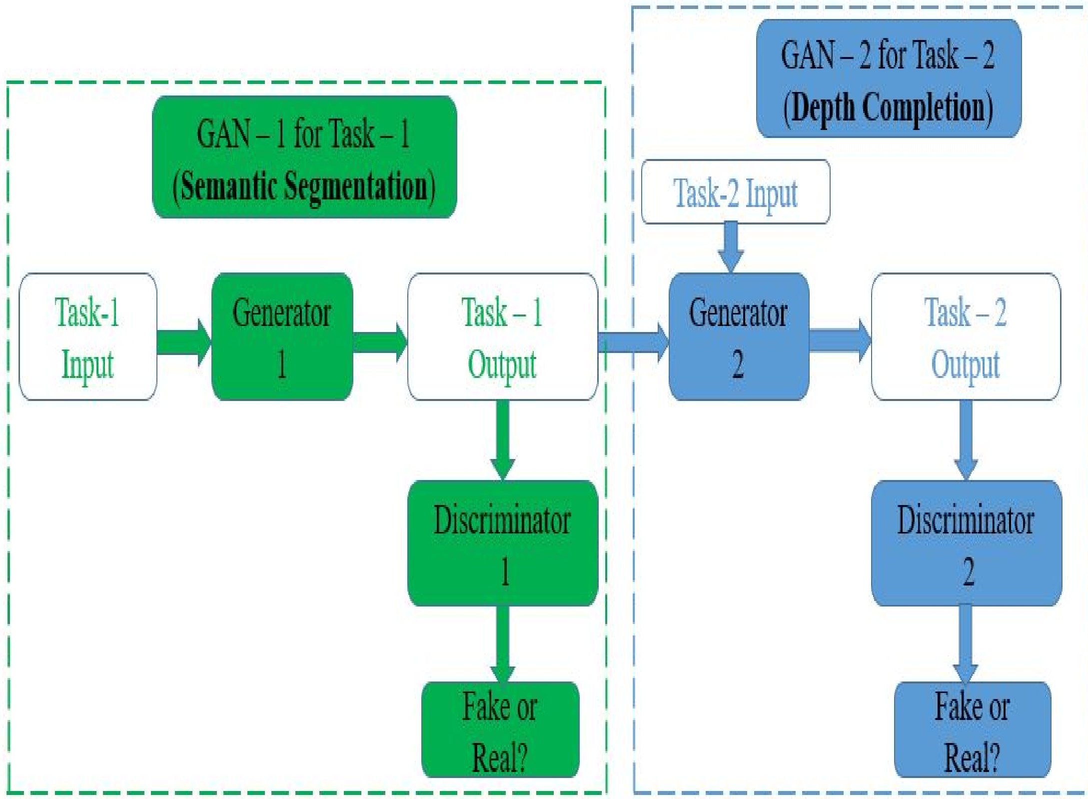


Figure 4.9 Multi Tasking GAN from [72]

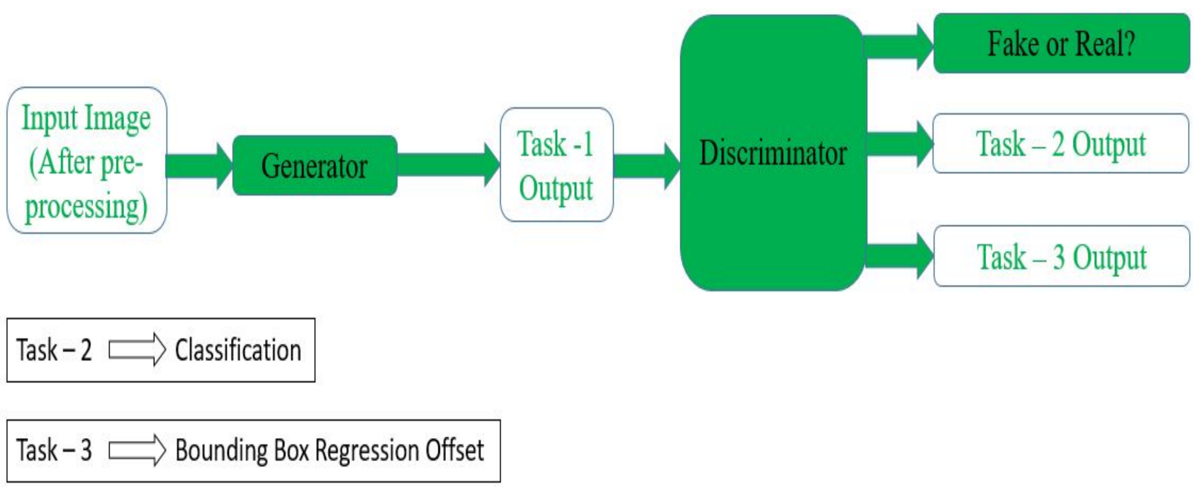


Figure 4.10 Diagram for MTGAN from [73]

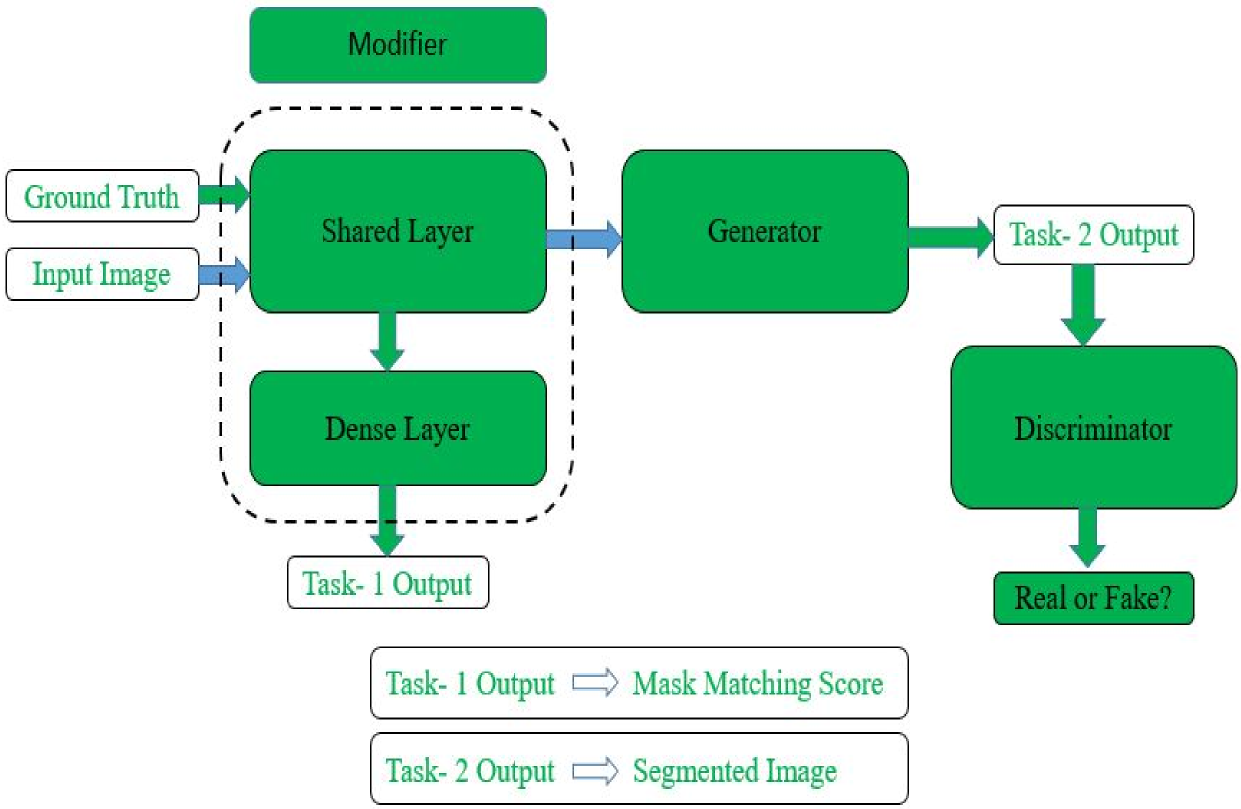


Figure 4.11 Task based Mod-CGAN implementation

Chapter Five

Performance Evaluation of the Proposed System

In this chapter, the impact of our proposed approach will be evaluated in terms of the baseline architecture. In this case, quantitative assessment based on the average dice score co-efficient will be determined in terms of comparison with the baseline architecture of CGAN named Pix2Pix. After that, the proposed approach will implemented along with its different versions and the results will be compared among themselves and also with the Pix2Pix framework. In line with this, performance of the imposed approach for different dataset will also be assessed. Also, the loss functions associated with both the generator and discriminator will be analyzed so that the impact of the modifier network can be assessed more deeply. After that, our proposed approach will be analyzed for a different GAN based network named SegNet-cGAN, which is very popular in the field of biomedical image segmentation task. Here, our proposed network, named as Mod-SegNet-cGAN, will be assessed in terms of both SegNet-cGAN as well as another state of the art approach named Extended U-Net. In the end, qualitative assessment based on all of the described network have been made.

5.1 Dataset Details

The research has been carried out generally on a publicly available dataset of only 163 US raw images and corresponding 163 skillfully annotated segmented labels collected by UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell (Spain) with a Siemens ACUSON Sequoia C512 System 17L5 HD Linear Array Transducer (8.5 MHz). Among them, randomly picked 125 images have been selected for training purpose and remaining 38 images have been selected for testing purpose. Assessment has been made on the basis of comparing average dice co-efficient of the test dataset. In this book, we have named this dataset as **Dataset I**. Along with this, in order to have better insight about our work, we have also used another dataset, acquired from [74]. This dataset in [74] contains breast ultrasound images of women aged between 25 and 75 years old. This dataset consists of 780 images with an average size of 500*500 pixels. The images are categorized into three classes named as benign, malignant, normal. It also includes ground truth annotation by expert radiologists. In this book, we have termed this dataset as **Dataset II**.

5.2 Simulation Environment

First of all, popular CGAN based implementation named Pix2Pix has been implemented for our ultrasound based biomedical dataset. As already described in Chapter 2, Pix2Pix is a encoder decoder based network and it has already been successfully implemented in literature for various applications in the field of natural images. So, its modified version on the basis of Multi Task Learning named Mod-CGAN has been implemented such that our proposed modification can be evaluated in line with its original framework. The hyper parameters used for Pix2Pix and Mod-CGAN models have been described as follows:

Features	**Pix2Pix**	**Mod-CGAN**
Nos. of sub network	02	03

Features	**Pix2Pix**	**Mod-CGAN**
Sub Network Name	Generator, Discriminator	Generator, Discriminator, Modifier
Nos. of Generator Layer used	19	16
Nos. of Modifier Layer used	No modifier network	04
Nos. of Discriminator Layer used	05	05
Learning rate for Generator and Discriminator	$2x10-04$	$2x10-04$
Learning rate for Modifier	No Modifier Network	$1x10-08$
Cross Entropy Utilized	Sigmoid	Sigmoid
Data Augmentation	No data augmentation	No data augmentation
Size of Train Dataset	125	125
Size of Test Dataset	38	38
Average epochs simulated	391	246

Table 5.1 Structural Difference between Pix2Pix and Mod-CGAN

It is to be noted from Table 5.1 that modifier learning rate has been made very low compared to that of generator as well as discriminator. It can be attributed to the fact that modifier network is a very small but effective addition from its part for the generator. So, it needs to be synchronized with the learning of comparatively large sized generator.

Otherwise, possibility here is that generator may be passing through some transition period whereas modifier has already ended its learning.

5.3 Analysis

5.3.1 Quantitative Assessment of Segmentation performance

With a view to assessing the model's performance quantitatively regarding the segmentation task, evaluation metrics in the form of Dice Score Co-efficient (DSC) has been utilized. Details about DSC have already been discussed in chapter 1. For the sake of comparison, whole dataset has been segregated into four folds on which both of the above described models have been assessed. In Table 5.2, quantitative assessment based on Average Dice Score Co-efficient have been highlighted for both of the models. Mod-CGAN is a kind of further enhancement of CGAN and also better utilization of available resources. This enhancement has been clearly indicated in the improvement of performance for Mod-CGAN.

Dice Score Co-efficient (DSC)		
Fold	**Pix2Pix**	**Mod-CGAN**
Fold-1	0.663496481	0.761219858
Fold-2	0.784143845	0.774587966
Fold-3	0.659522533	0.696189301
Fold-4	0.776714678	0.802293774
Average DSC	**0.720969384**	**0.758572725**
Improvement	**-**	**5.215663974%**

Table 5.2 Quantitative Assessment on the dataset

5.3.2 Loss Function Analysis

In this section, loss function corresponding with the three networks for our architecture will be shown graphically. Then, it will be compared with the loss functions of CGAN architecture.

Figure 5.1 shows loss function against no. of iterations associated with three different networks of Mod-CGAN. Loss values have been shown for no. of iterations up to the point of convergence. For reaching to the convergence of our architecture, we have to iterate the

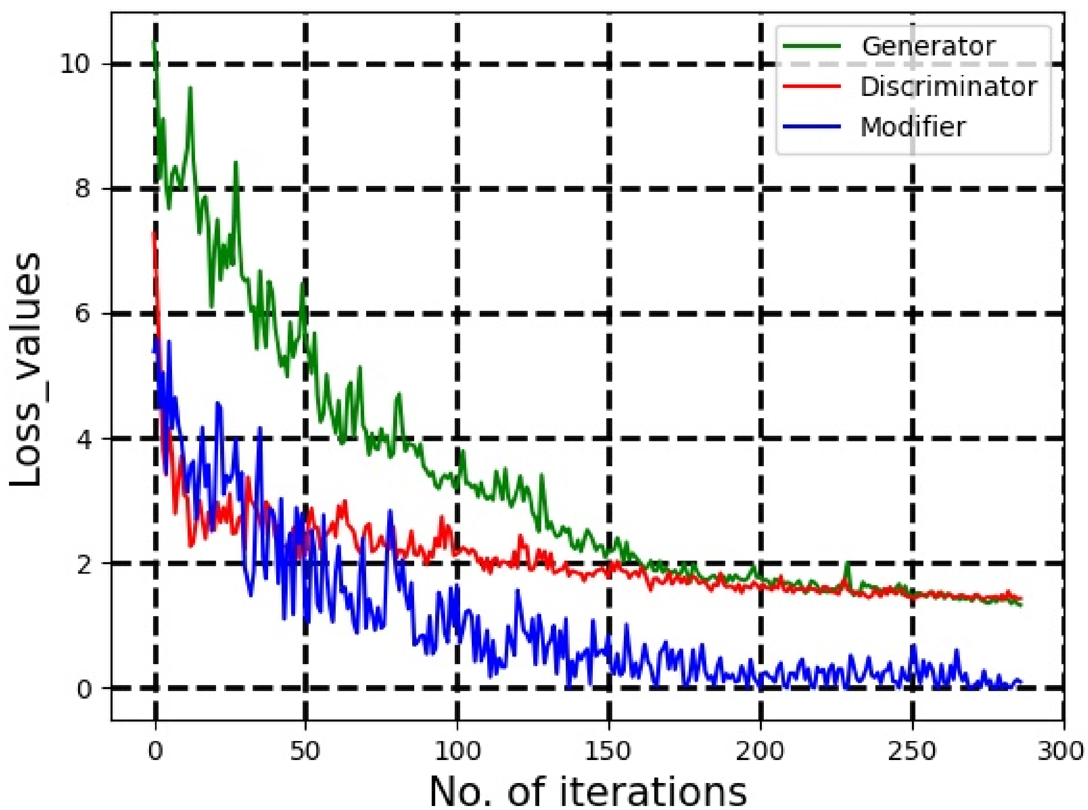


Figure 5.1 Loss functions regarding three networks of Mod-CGAN

model around 256 times for this specific fold. But for the issue of checking stability in Mod-CGAN, we have run our model for around 288 times. In these extra 32 epochs, although generator improves its individual performance, but still its generalization ability remains fixed.

In Figure 5.2, the loss function regarding two networks of CGAN has been shown graphically. For reaching to the convergence of Conditional GAN architecture, we have to impose

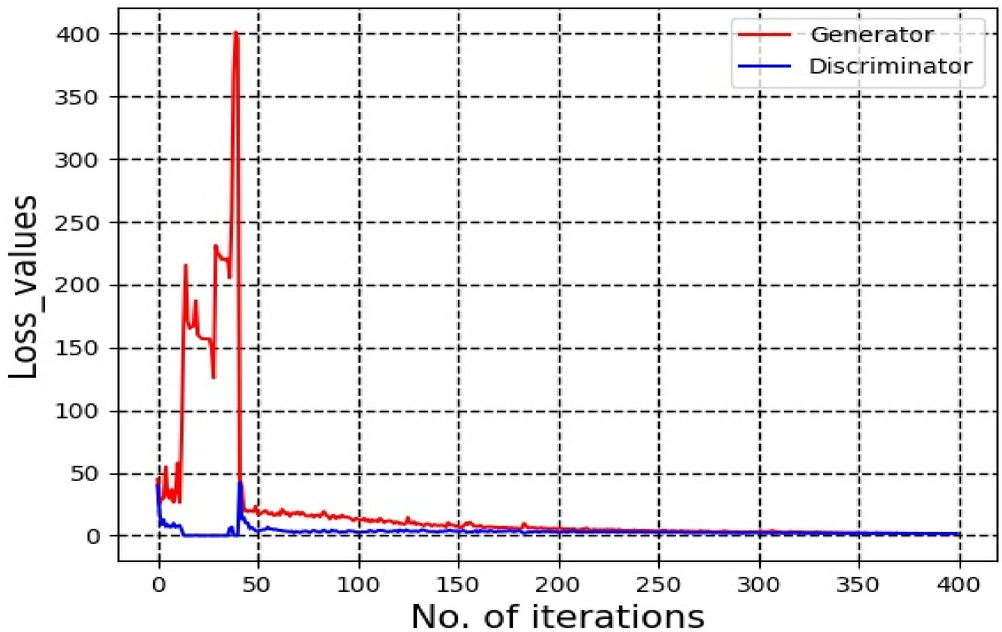


Figure 5.2 Loss functions regarding two networks of CGAN

the architecture on the training dataset for around 400 times. Also, it can be clearly noted that up to some iterations, generator behaves very randomly and therefore the loss value is very high. After then, generator starts to track the path to convergence. Whatever, with the introduction of our very simple modifier network, generator, from the very beginning, remains very much focussed on the assigned task. Moreover, in the later phase, the speed of generator's tracking to the convergence is much higher than the speed of CGAN as our model reaches to its point of convergence after 256 epochs comparing to the 400 epochs of CGAN. The case of generator and discriminator for both the architectures are shown in figure 5.3 and figure 5.4 respectively for better understanding. So, it is evident from figure 5.4 that level of discriminator for Mod-CGAN has been achieved by the level of discriminator for

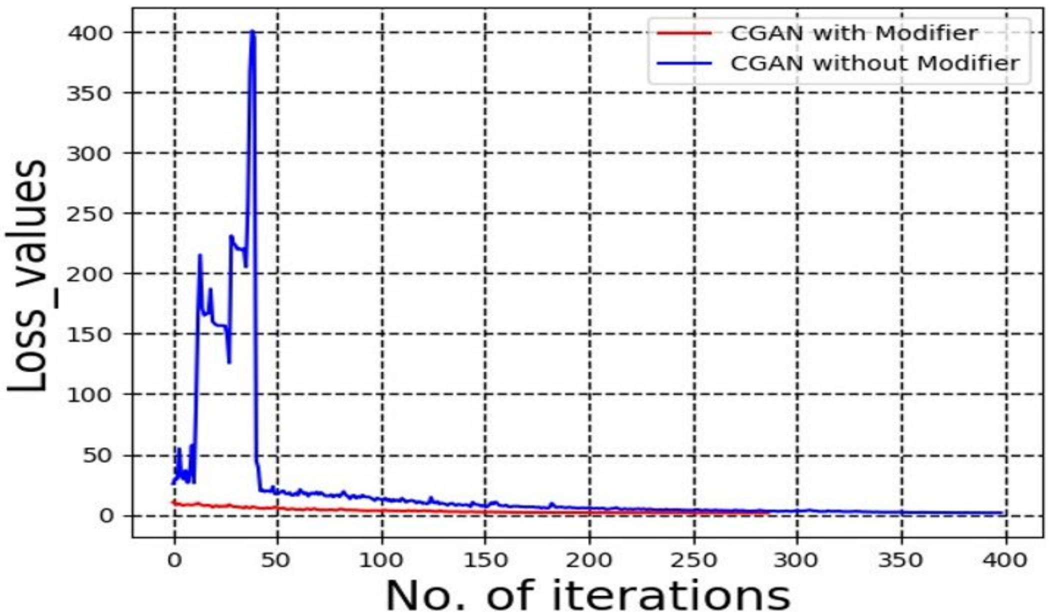


Figure 5.3 Generator loss functions for two models: CGAN and Mod-CGAN

CGAN after around 100 epochs later. Correspondingly, the case is same for generator. Also, from figure 5.3, it can be inferred about the much higher convergence speed for generator of Mod-CGAN model than the CGAN model.

5.3.3 Different versions of Mod-CGAN

Our method utilizes the scope of multi tasking learning. As depicted from [75], there exists two types of hidden layers in a network, i.e. the shared layers and the task specific layers. The shared layers learn the intrinsic low level representation of the data, which are general among all the tasks, while the task specific layers learn the network's task specific parameters that map the learned latent representations from the previous shared layers to the task specific output layers. In short, as we go deep into the hidden layers, the correspondence of its features will be more and more task oriented. As, MTL deals with the issue of joint feature learning which is common among all the tasks, so if we bring more encoder layers within

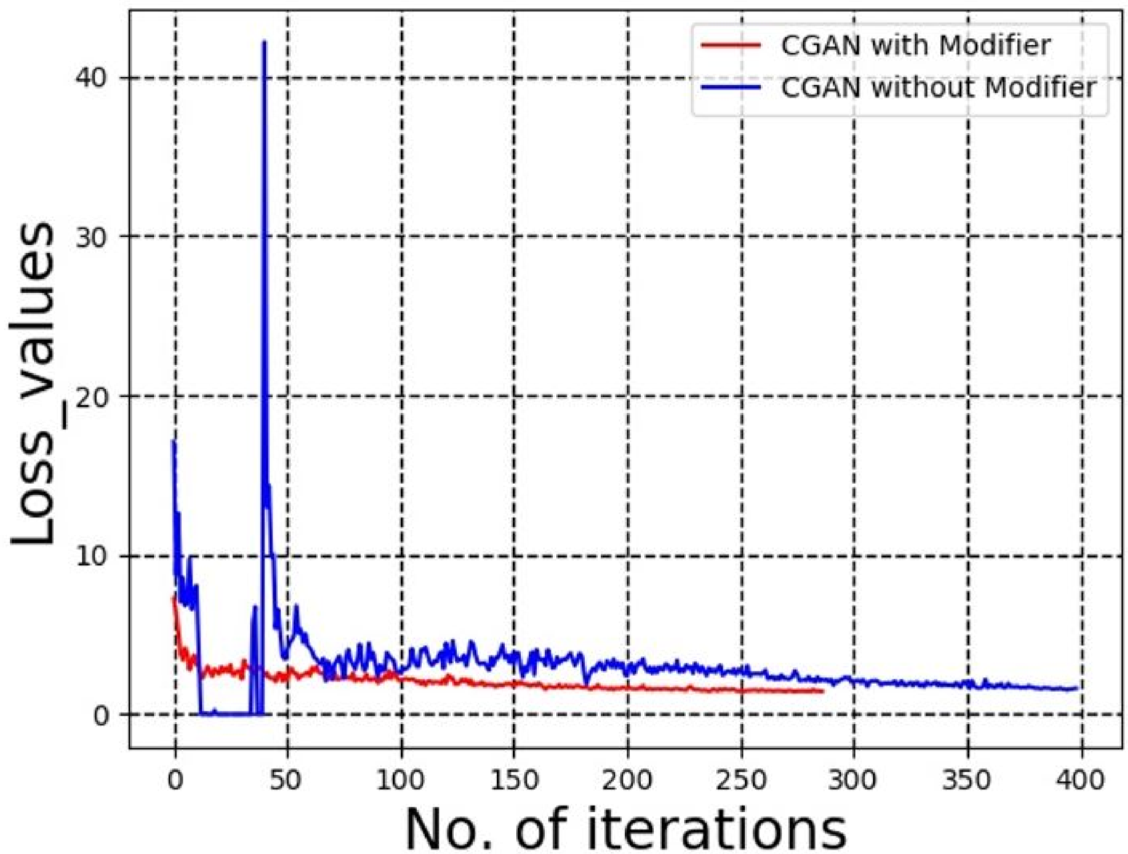


Figure 5.4 Discriminator loss functions for two models: CGAN and Mod-CGAN

the modifier architecture, the possibility will be to enter into the task specific layers, due to which our method can lose its generality, thereby leading to overfitting. In Table 5.3, we have shown average DSC on a specific fold of our dataset for different versions of Mod-CGAN. In the table mentioned above, Mod-CGAN refers to the same as our proposed method. (Mod-CGAN) - 1 refers to Mod-CGAN having one less modifier layer and correspondingly that layer will be included in the generator network. Similarly, (Mod-CGAN) + 1 refers to Mod-CGAN having one more modifier layer and a less generator layer than our proposed Mod-CGAN. (Mod-CGAN) + 2 has also been described in this way.

After analyzing table 5.2 and table 5.3, it is clearly evident that Mod-CGAN, (Mod-CGAN) - 1, (Mod-CGAN) +1 perform better than original core GAN network, Pix2Pix.

Dice Score Co-efficient (DSC)			
(Mod-CGAN) - 1	**Mod-CGAN**	**(Mod-CGAN) + 1**	**(Mod-CGAN) + 2**
0.72951096	**0.77716871**	**0.76105495**	**0.67999989**

Table 5.3 DSC associated with different variants of Mod-CGAN

From (Mod-CGAN) + 2, network degrades in generalization capability, which refers to the fact that modifier network has been deep into task specific layers rather than being confined into shared hidden layers. Figure 5.5 shows some visual depiction for different versions of Mod-CGAN. Apart from the assessment made in Table 5.3, it is also of interest to have more insight through loss function of its different parts. For the case of Generator loss function as depicted in Figure 5.6, although the Generator loss values for (Mod-CGAN) +2 is less than other Mod-CGAN variant, its performance is worst among the four Mod-CGAN variants as evident from Table 5.2. Here another issue is obvious that convergence is achieved quite early for (Mod-CGAN) +2 than other Mod-CGAN variants. Therefore, with the introduction of more deeper layers into the modifier part, model starts to lose its generalization capability. As apparent from the distinction between shared and task specific layers, we have dived deep into task specific layers in (Mod-CGAN) +2 version. Now for the case of Discriminator loss function as shown in Figure 5.7, Discriminator is more optimized in Mod-CGAN (our best performed model) compared with other Mod-CGAN variants, thereby leading the Generator to optimize itself through more adversaries. Here, one thing needs to be mentioned that upon iterating (Mod-CGAN) +2 for much more epochs, its generalization capability degrades quite rapidly, although its generator loss reduces then. From Figure 5.8, there is nothing significant to interpret as modifier acts as a performance booster as well as an assistance for generator.

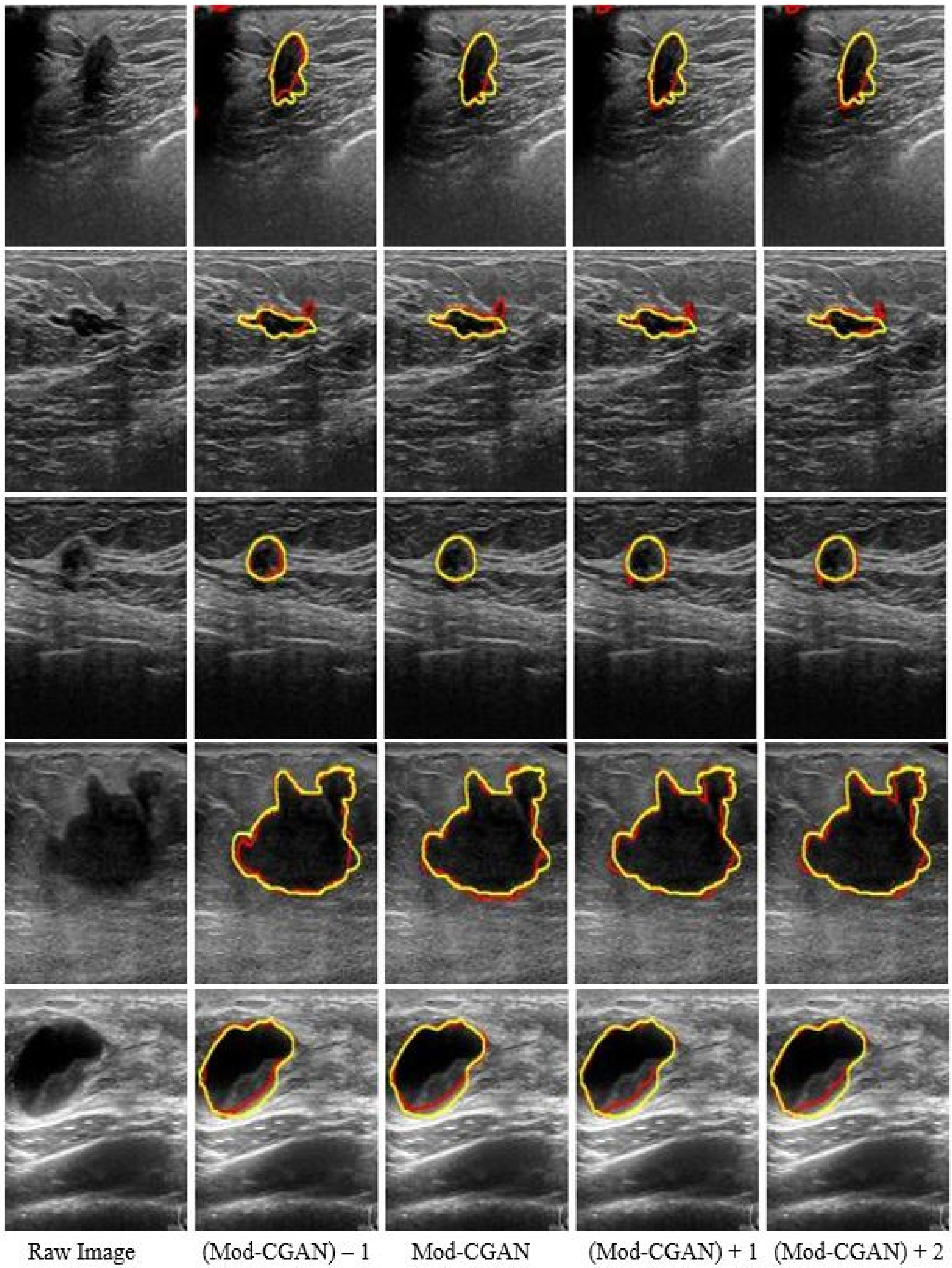
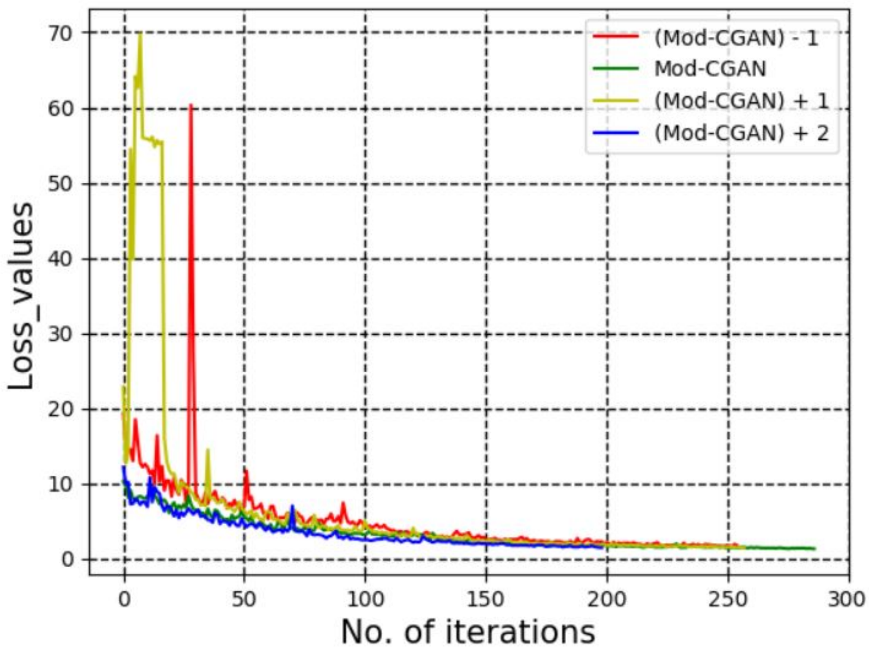
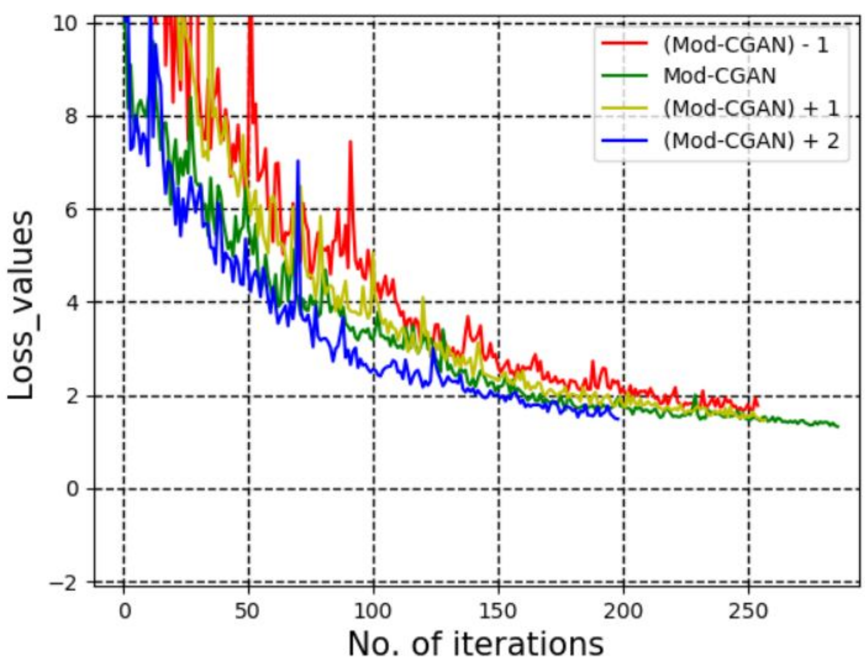


Figure 5.5 Depiction for different versions of Mod-CGAN

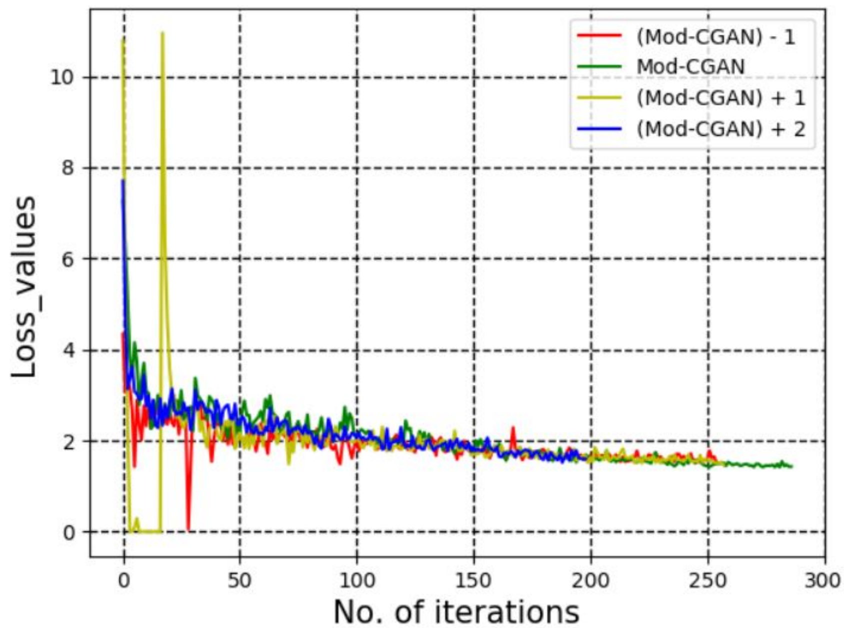


(a) Complete Figure

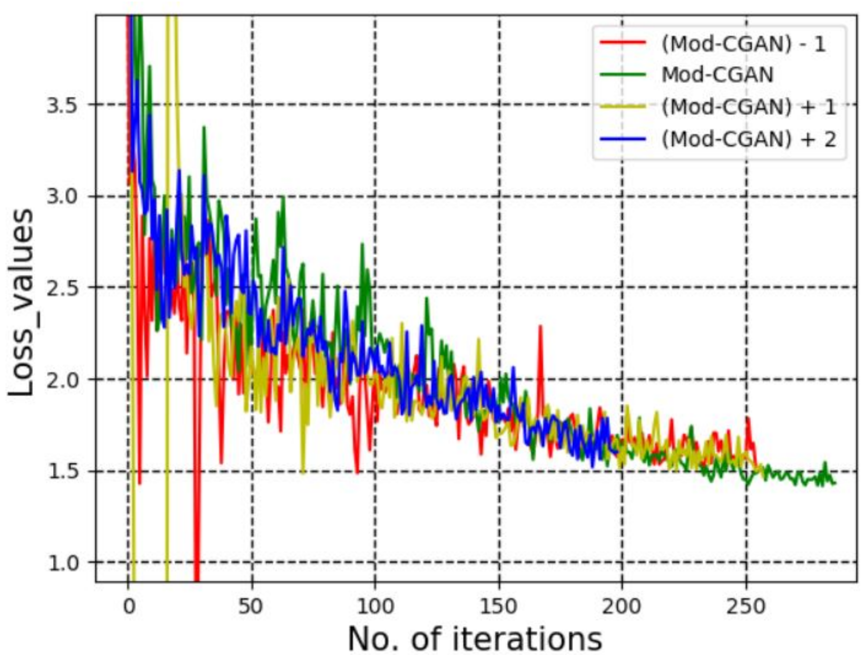


(b) Zoom Version

Figure 5.6 Generator loss function (Original and Zoom Version) for different versions of Mod-CGAN



(a) Complete Figure



(b) Zoom Version

Figure 5.7 Discriminator loss function (Original and Zoom Version) for different versions of Mod-CGAN

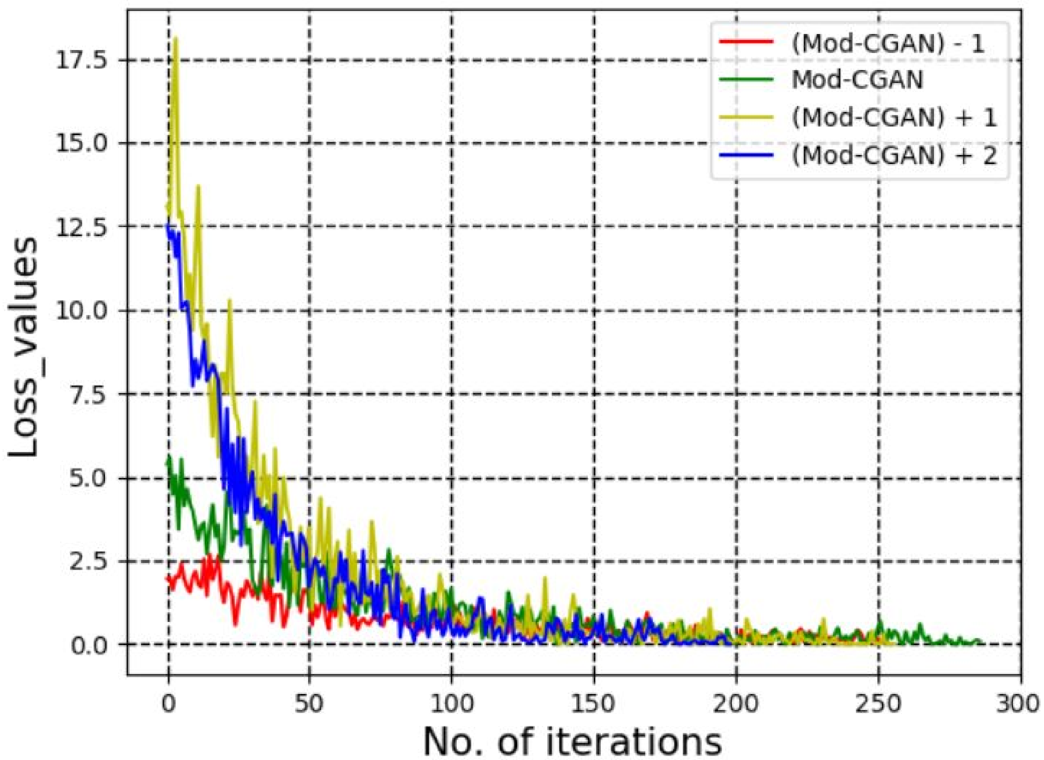


Figure 5.8 Modifier loss function for different versions of Mod-CGAN

5.3.4 Performance Evaluation for Dataset II

In this section, we will consider Dataset II for the evaluation of our proposed approach. As we are interested in the segmentation of breast lesion from ultrasound images, so we only take into consideration the benign and malignant lesions. Based on this, size of the train dataset is 504 images and size of the test dataset is 126 images. Both of the above described models have been simulated for this dataset and average dice score co-efficient has been evaluated as shown in Table 5.4.

Dice Score Co-efficient (DSC)	
Pix2Pix	**Mod-CGAN**
0.74056696	0.75596155
Improvement	**2.07875733%**

Table 5.4 Quantitative Assessment on Dataset II

5.4 Performance evaluation of Modifier imposition for different CGAN architecture

In the previous subsection, imposition of modifier network on a basic CGAN framework has been assessed in the form of Mod-CGAN architecture. In this section, another model named SegNet-cGAN has been analyzed. SegNet-cGAN has been successfully implemented for X-Ray based breast mass segmentation in 2018 [68] and also for Mammogram based breast tumor segmentation in 2020 [69]. Based on this, this model has been imposed for Ultrasound based breast tumor segmentation which brings a huge improvement compared to that of Pix2Pix. Model assessment through average dice score coefficient based on the four fold on Dataset I as previously described has been shown in Table 5.5.

Dice Score Co-efficient (DSC)		
Fold	**Pix2Pix**	**SegNet-cGAN**
Fold-1	0.6635	0.8164
Fold-2	0.7841	0.8236
Fold-3	0.6595	0.7431
Fold-4	0.7767	0.8017
Average DSC	**0.7210**	**0.7962**
Improvement	**-**	**10.4299%**

Table 5.5 Quantitative Assessment between Pix2Pix and SegNet-cGAN

As results have been improved significantly compared to Pix2Pix network, so our multi tasking based modifier implementation can be imposed on this SegNet-cGAN approach, which can be descibed as Mod-SegNet-cGAN. Comparison between architectural features has been given in table 5.6.

Features	**SegNet-cGAN**	**Mod-SegNet-cGAN**
Total nos. of Generator Layer	42	39
Total nos. of Modifier Layer	No modifier network	04
Total nos. of Discriminator Layer	14	14
Cross Entropy Utilized	Sigmoid	Sigmoid
Data Augmentation	No data augmentation	No data augmentation

Table 5.6 Structural Features of SegNet-cGAN and Mod-SegNet-cGAN

From Table 5.6, it can easily be stated that SegNet-cGAN is a large network compared to Pix2Pix network. Now, in Table 5.7, performance of Mod-SegNet-cGAN has been assessed with respect to SegNet-cGAN as well as Extended U-Net approach [61] as a different upgraded form of U-Net architecture.

Dice Score Co-efficient (DSC)			
Fold	**Extended U-Net**	**SegNet-cGAN**	**Mod-SegNet-cGAN**
Fold-1	0.788873235	0.816445409	0.823337554
Fold-2	0.656127464	0.823639439	0.838176944
Fold-3	0.744901166	0.743076897	0.750324451
Fold-4	0.572751655	0.801700179	0.798747811
Average DSC	**0.69066338**	**0.796215481**	**0.80264669**
Improvement	**-**	**15.28271283%**	**16.21387686%**

Table 5.7 Quantitative Assessment on the dataset

Also it can be stated that improvement for our proposed approach from SegNet-cGAN is

0.8077%. So, the improvement has been made, but it is small compared to the case between Pix2Pix and Mod-CGAN. It may be due to the fact that SegNet-cGAN is a very large network in comparison with Pix2Pix. So, there is a possibility that network may reach to its limit with this dataset. As a result, the imposition of this small, performance boosting modifier network into the CGAN implementation improves the result for both the small and large networks, but not in a very significant way for large network.

5.5 Qualitative interpretation on Dataset I

Figure 5.9 depicts the qualitative assessment on Dataset I based on different cases like small sized lesions, complex lesion boundary lines, lesion surrounding fuzzy contour lines etc. This figure also infers about the superior performance for the modifier based implementation.

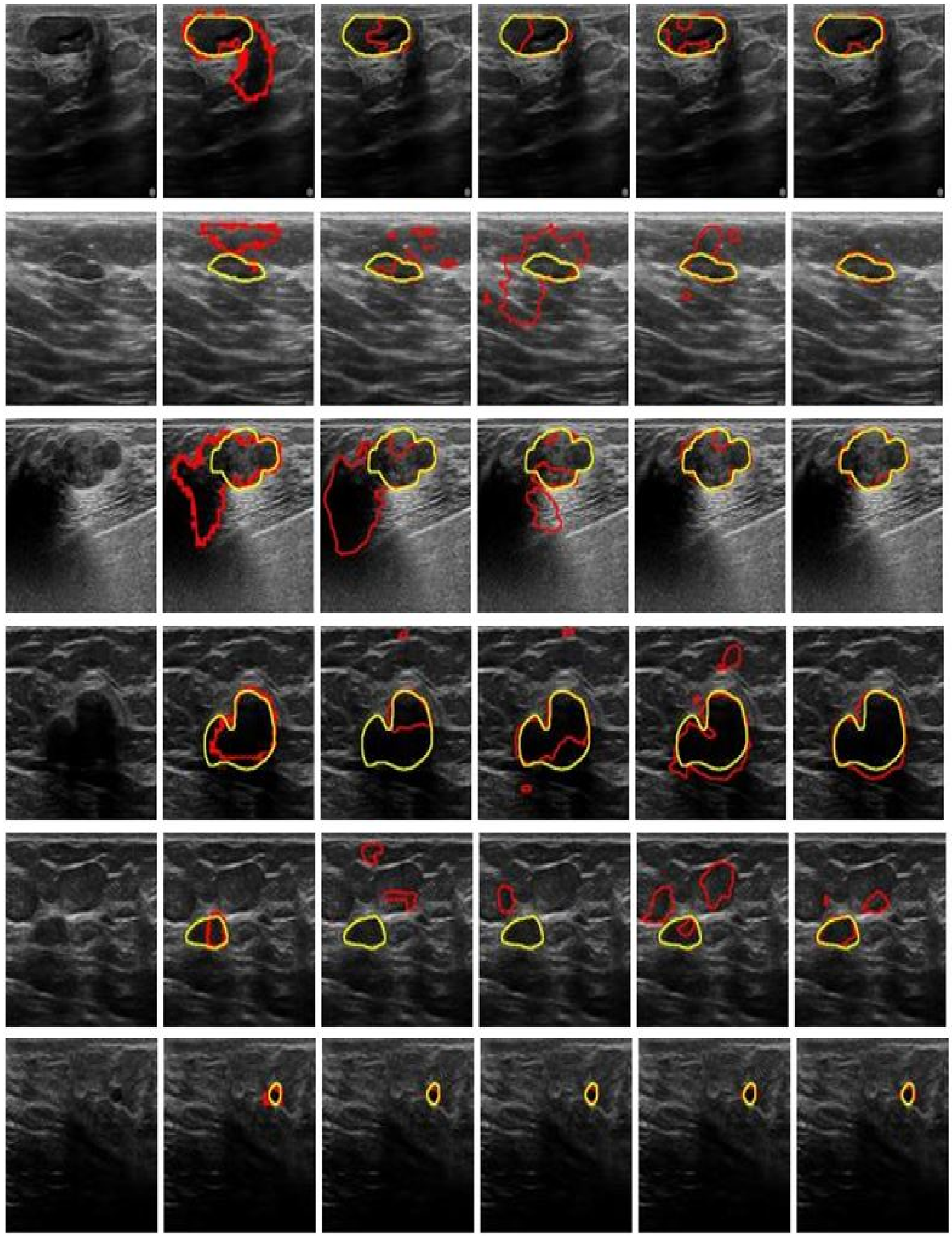


Figure 5.9 Qualitative Assessment on dataset I

Chapter Six

Conclusion

This thesis work deals with the segmentation of breast lesions from Ultrasound Images using of Conditional Generative Adversarial Network. Our proposed system can identify, localize as well as segment the breast lesions from the fuzzy and mostly confusing Ultrasound images. Here, the system has been designed with a view to utilizing the scarcely available annotated dataset. In general, conditional GAN has the luxury to be trained in a supervised way through the introduction of both raw images and labeled images in the loss function. Our proposed system is also a supervised one but by the use of multi tasking learning, it can lead the generator to the path of convergence much easily. In this way, convergence as well as stability of conditional GAN has been ensured. Also, our work tends to perform within the territory of a very small dataset without the risk of overfitting or underfitting and data augmentation has also not been incorporated here. In order to evaluate its performance on the specific task, both the architectures of U-Net and conditional GAN have been investigated which ensures the fact that our proposed system can surpass them and behaves quite well in pretty critical condition. In addition to this, significance of our imposition has been analyzed for different dataset. The performance improvement of our proposed imposition for Dataset I are as follows:

- Mod-CGAN performs **5.22%** better than Pix2Pix (CGAN).
- Mod-SegNet-cGAN performs **16.21%** better than Extended U-net.

- Mod-SegNet-cGAN performs **0.81%** better than SegNet-cGAN.

whereas, performance improvement of our proposed imposition for Dataset II is as follows:

- Mod-CGAN performs **2.08%** better than Pix2Pix (CGAN).

In the conclusion, we can include some of the facts that may possess the scope of further advancement as per following:

★ Our proposed modifier network includes only one dense layer after the shared network. As we dig deep from Mod-CGAN to (Mod-CGAN) + 1 as well as (Mod-CGAN) + 2, performance degrades in terms of generalization capability. It is evident that as the no. of latent representation layers increase in the modifier network, the no. of features as well as neurons rise quite extremely. As a result, conversion from such a high number of features/ neurons to a single value in one single dense layer may insert the possibility of losing general data. So, here the question arises whether the inclusion of additional dense layers in the case of higher versions of Mod-CGAN is beneficial or not. Also, the performance of different hyper parameters at different values needs to be evaluated, as it might cast some more insight into the convergence and stability of the overall network.

★ Our proposed system is only assessed on the task of biomedical image segmentation. As we know that deep learning approaches have been incorporated into biomedical imaging task and it has thereby changed the view of medical imaging analysis, so we can anticipate that it will obviously have a huge influence on computer vision based real time tasks. The possibility here is that it can replace for example, the so-called state of the art object recognition system.

★ Our work has been implemented without the exercise of any data augmentation, such that work can be analyzed for the core network. Although, data augmentation can not assimilate the natural deformation of the internal body organs, still exercise of data augmentation for enlarging the dataset could be a point of case study for our proposition.

- ★ Our proposed architecture is more of its kind of supervised training, opening the door for more research to migrate towards the kind of unsupervised training.

- ★ In our datasets, we have single breast lesion to detect in each of its individual image. As, this architecture is not trained for the rare case of more than one lesion detectable in our original raw image, we should exaggerate this model's evaluation towards determining lesions in all the possible places of single uniform image. In this way, it also can be guaranteed whether our method is properly generalized or not.

- ★ Our proposed imposition has been utilized for both small and large CGAN implementation. Although, performance has been improved significantly for small network, performance improvement is no that much large for very big network. So, further analysis can be made regarding performance improvement for large network.

- ★ In the case of GAN network training, there has been provision of implementing different cost functions, none of which still are not totally optimized for all the datasets. One cost function may perform superior for one dataset, whereas another cost function performs better than other for other dataset. Research has been ongoing still to find the most optimized one which can outperform every other cost functions for every dataset. A glimpse of still available cost function for GAN training purpose are covered in chapter 2. It also gives the scope for further extension of our work.

REFERENCES

[1] Cheng, H. D., Shan, J., Ju, W., Guo, Y., and Zhang, L., "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognition*, vol. 43, no.1, pp. 299-317, 2010.

[2] Singh, V.K., Rashwan, H.A., Romani, S., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Arenas, M., Arquez, M., Puig, D., and Torrents-Barrena J., “Breast Tumor Segmentation and Shape Classification in Mammograms using Generative Adversarial and Convolutional Neural Network,” *Expert Systems with Applications*, vol. 139, p. 112855, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417419305573.

[3] Pan,H.- B.,"The Role of Breast Ultrasound in Early Cancer Detection," *Journal of Medical Ultrasound*, vol. 24, no. 4, pp. 138-141, 2016.

[4] Lee, C. I., Chen, L. E., and Elmore, J. G., .,"Risk-Based Breast Cancer Screening: Implications of Breast Density," *The Medical clinics of North America*, vol. 101, no. 4, pp. 725-741, 2017.

[5] Nazari, S. S., and Mukherjee, P., "An overview of mammographic density and its association with breast cancer," *Breast Cancer (Tokyo, Japan)*, vol. 25, no. 3, pp. 259-267, 2018.

[6] Huynh, B., Drukker, K., and Giger, M., "Mo-de-207b-06: Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks," *Medical Physics*, vol. 43, no. 6, pp. 3705-3705, 2016.

[7] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A.A., "Image-to-image translation with conditional adversarial networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125-1134, 2017.

[8] Shan, J., Cheng, H.D., and Wang, Y., "Completely Automated Segmentation Approach for Breast Ultrasound Images using Multiple-Domain Features," *Ultrasound in Medicine Biology*, vol. 38, no. 2, pp. 262-275, 2012.

[9] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234-241, 2015.

[10] Drukker, K., Giger, M.L., Horsch, K., Kupinski, M.A., Vyborny, C.J., and Mendelson, E.B., "Computerized lesion detection on breast ultrasound," *Medical Physics*, vol. 29, no. 7, pp. 1438-1446, 2002.

[11] Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W., "Photo-realistic single image super-resolution using a Generative Adversarial Network," *in Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, pp. 4681-4690, 2017.

[12] Hitawala, S., "Comparative Study on Generative Adversarial Networks," arXiv preprint arXiv: 1801.04271, 2018.

[13] Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. and Lanczi, L., "The multimodal brain tumor image segmentation benchmark (BRATS)", *in IEEE transactions on medical imaging*, no. 34(10), pp.1993-2024, 2014.

[14] Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., De Bresser, J., Alansary, A., De Bruijne, M., Carass, A., El-Baz, A. and Jog, A., "MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans", *in Computational intelligence and neuroscience*, 2015.

[15] Išgum, I., Benders, M.J., Avants, B., Cardoso, M.J., Counsell, S.J., Gomez, E.F., Gui, L., Hűppi, P.S., Kersbergen, K.J., Makropoulos, A. and Melbourne, A., "Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge", *in Medical image analysis*, no. 20(1), pp.135-151, 2015.

[16] Wang, L., Nie, D., Li, G., Puybareau, É., Dolz, J., Zhang, Q., Wang, F., Xia, J., Wu, Z., Chen, J.W. and Thung, K.H., "Benchmark on automatic six-month-old infant brain segmentation algorithms: the iSeg-2017 challenge", *in IEEE transactions on medical imaging*, no. 38(9), pp.2219-2230, 2019.

[17] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[18] Salakhutdinov, R. and Hinton, G., "Deep Boltzmann Machines" *in Artificial intelligence and statistics*, pp. 448-455, 2009.

[19] Le, Q.V., "Building High-Level Features Using Large-Scale Unsupervised Learning," *in IEEE International Conference on acoustics, speech and signal processing*, pp. 8595-8598, 2013.

[20] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E. and Darrell, T., "DeCAF — A Deep Convolutional Activation Feature for Generic Visual Recognition," *in International conference on machine learning*, pp. 647-655, 2014.

[21] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M., "Playing Atari with Deep Reinforcement Learning", arXiv preprint arXiv:1312.5602, 2016.

[22] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions", *In Proceedings of*

the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015.

[23] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition", *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[24] Veit, A, Wilber, M.J., and Belongie, S., "Residual networks behave like ensembles of relatively shallow networks", *In Advances in neural information processing systems*, pp. 550-558, 2016.

[25] Kobler, E., Klatzer, T., Hammernik, K. and Pock, T. "Variational networks: connecting variational methods and deep learning", *In German conference on pattern recognition*, pp. 281-293, 2017.

[26] Sak, H., Senior, A.W. and Beaufays, F., "Long short-term memory recurrent neural network architectures for large scale acoustic modeling", 2014.

[27] Chung, J., Gulcehre, C., Cho, K. and Bengio, Y., "Empirical evaluation of gated recurrent neural networks on sequence modeling", arXiv preprint arXiv: 1412.3555, 2014.

[28] Hinton, G.E., and Salakhutdinov, R.R., "Reducing the dimensionality of data with neural networks", *in science*, no. 313(5786), pp.504-507, 2006.

[29] Goodfellow, I., Bengio, Y. and Courville, A., "Deep learning", *in MIT press*, 2016.

[30] Bank, D., Koenigstein, N. and Giryes, R., "Autoencoders", arXiv preprint arXiv: 1412.3555, 2020.

[31] Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z.B., and Swami, A., "The limitations of deep learning in adversarial settings", *In IEEE European symposium on security and privacy (EuroSP)*, pp. 372-387, 2016.

[32] Goodfellow, I.J., Shlens, J., and Szegedy, C., “Explaining and harnessing adversarial examples”, arXiv preprint arXiv:1412.6572, 2014.

[33] Madry, A., Makelov, A., Schmidt, L., Tsipras, D. and Vladu, A., “Towards deep learning models resistant to adversarial attacks”, arXiv preprint arXiv:1706.06083, 2017.

[34] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., “Generative adversarial nets”, *In Advances in neural information processing systems*, (pp. 2672-2680), 2014.

[35] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D., "StackGAN++: Realistic image synthesis with stacked generative adversarial networks," *in IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 41, no. 8, pp. 1947-1962, 2018.

[36] Gatys, L.A., Ecker, A.S. and Bethge, M., “Image style transfer using convolutional neural networks”, *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414-2423, 2016.

[37] Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W., "Photo-realistic single image super-resolution using a Generative Adversarial Network," *in Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, pp. 4681-4690, 2017.

[38] Nagarajan, V. and Kolter, J.Z., “Gradient descent GAN optimization is locally stable”, *In Advances in neural information processing systems*, pp. 5585-5595, 2017.

[39] Roth, K., Lucchi, A., Nowozin, S. and Hofmann, T., “Stabilizing training of generative adversarial networks through regularization”, *In Advances in neural information processing systems*, pp. 2018-2028, 2017.

[40] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X., “Improved techniques for training gans”, *In Advances in neural information processing systems*, pp. 2234-2242, 2016.

[41] Lucic, M., Kurach, K., Michalski, M., Gelly, S. and Bousquet, O., “Are gans created equal? a large-scale study”, *In Advances in neural information processing systems*, pp. 700-709, 2018.

[42] Mirza, M., and Osindero, S., “Conditional generative adversarial nets”, arXiv preprint arXiv:1411.1784, 2014.

[43] Li, C. and Wand, M., “Precomputed real-time texture synthesis with markovian generative adversarial networks”, *In European conference on computer vision*, pp. 702-716, 2016.

[44] Pathak,D., Krahenbuhl,P.,Donahue,J., Darrell, T., and A. Efros, A., "Context encoders: Feature learning by inpainting", *CVPR*, 2016.

[45] Van Ginneken, B., Frangi, A.F., Staal, J.J., ter Haar Romeny, B.M., and Viergever, M.A., “Active shape model segmentation with optimal features”, *In IEEE transactions on medical imaging*, no. 21(8), pp.924-933, 2002.

[46] Lo, S.C., Lou, S.L., Lin, J.S., Freedman, M.T., Chien, M.V. and Mun, S.K., “Artificial convolution neural network techniques and applications for lung nodule detection”, *in IEEE transactions on medical imaging*, no. 14(4), pp.711-718, 1995.

[47] Krizhevsky, A., Sutskever, I. and Hinton, G.E., “Imagenet classification with deep convolutional neural networks”, *In Advances in neural information processing systems*, pp. 1097-1105, 2012.

[48] Shen, D., Wu, G. and Suk, H.I., “Deep learning in medical image analysis”, *Annual review of biomedical engineering*, no. 19, pp. 221-248, 2017.

[49] Bengio, Y., Courville, A. and Vincent, P., “Representation learning: A review and new perspectives”, *in IEEE transactions on pattern analysis and machine intelligence*, no. 35(8), pp.1798-1828, 2013.

[50] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B. and Sánchez, C.I., “A survey on deep learning in medical image analysis”, *Medical image analysis*, no. 42, pp.60-88, 2017.

[51] A. Skourt,B., E. Hassani,A. and Majda,A., "Lung CT Image Segmentation Using Deep Neural Networks", *Procedia Computer Science*, vol. 127, pp. 109-113, 2018.

[52] Fennema-Notestine, C., Ozyurt, I.B., Clark, C.P., Morris, S., Bischoff-Grethe, A., Bondi, M.W., Jernigan, T.L., Fischl, B., Segonne, F., Shattuck, D.W. and Leahy, R.M., “Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location”, *Human brain mapping*, no. 27(2), pp.99-113, 2006.

[53] Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M. and Biller, A., “Deep MRI brain extraction: A 3D convolutional neural network for skull stripping”, *in NeuroImage*, no. 129, pp.460-469, 2016.

[54] Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S. and Shen, D., “Deep convolutional neural networks for multi-modality isointense infant brain image segmentation”, *in NeuroImage*, no. 108, pp.214-224, 2015.

[55] de Brebisson, A. and Montana, G., “Deep neural networks for anatomical brain segmentation”, *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 20-28, 2015.

[56] Moeskops, P., Viergever, M.A., Mendrik, A.M., De Vries, L.S., Benders, M.J. and Išgum, I., “Automatic segmentation of MR brain images with a convolutional neural

network", *in IEEE transactions on medical imaging*, no. 35(5), pp.1252-1261, 2016.

[57] Ciresan, D., Giusti, A., Gambardella, L.M. and Schmidhuber, J., "Deep neural networks segment neuronal membranes in electron microscopy images", *In Advances in neural information processing systems*, pp. 2843-2851, 2012.

[58] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S. and Pal, C., "The importance of skip connections in biomedical image segmentation", *In Deep Learning and Data Labeling for Medical Applications*, pp. 179-187, 2016.

[59] Milletari, F., Navab, N. and Ahmadi, S.A., "V-net: Fully convolutional neural networks for volumetric medical image segmentation", *In Fourth International Conference on 3D Vision (3DV)*, pp. 565-571, 2016.

[60] Yi, X., Walia, E. and Babyn, P., "Generative adversarial network in medical imaging: A review", *in Medical image analysis*, p.101552, 2019.

[61] Guo, Y., Duan, X., Wang, C. and Guo, H., "Segmentation and recognition of breast ultrasound images based on an expanded U-Net", *Plos one*, no.16(6), p.e0253202, 2021.

[62] Yang, D., Xu, D., Zhou, S.K., Georgescu, B., Chen, M., Grbic, S., Metaxas, D. and Comaniciu, D., "Automatic liver segmentation using an adversarial image-to-image network", *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 507-515, 2017.

[63] Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D. and Glocker, B., "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks", *In International conference on information processing in medical imaging*, pp. 597-609, 2017.

[64] Dou, Q., Ouyang, C., Chen, C., Chen, H. and Heng, P.A., "Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with

adversarial loss”, arXiv preprint arXiv:1804.10916, 2018.

[65] Xue, Y., Xu, T., Zhang, H., Long, L.R. and Huang, X., “Segan: Adversarial network with multi-scale L1 loss for medical image segmentation”, *in Neuroinformatics*, no. 16(3-4), pp.383-392, 2018.

[66] Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P. and Chen, D.Z., “Deep adversarial networks for biomedical image segmentation utilizing unannotated images”, *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 408-416, 2017.

[67] Zhu, W., Xiang, X., Tran, T.D., Hager, G.D. and Xie, X., “Adversarial deep structured nets for mass segmentation from mammograms”, *In IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (pp. 847-850), 2018.

[68] Singh, V.K., Romani, S., Rashwan, H.A., Akram, F., Pandey, N., Sarker, M., Kamal, M., Abdulwahab, S., Torrents-Barrena, J., Saleh, A. and Arquez, M., "Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification", *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 833-840, 2018.

[69] Singh, V.K., Rashwan, H.A., Romani, S., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Arenas, M., Arquez, M., Puig, D. and Torrents-Barrena, J., "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network", *Expert Systems with Applications*, vol. 139, p.112855, 2020.

[70] Noble, J.A. and Boukerroui, D., “Ultrasound image segmentation: a survey”, *in IEEE Transactions on medical imaging*, no. 25(8), pp. 987-1010, 2006.

[71] Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D. and Wang, T., "Deep learning in medical ultrasound analysis: a review", *Engineering*, 2019.

[72] Zhang, C., Tang, Y., Zhao, C., Sun, Q., Ye, Z. and Kurths, J., "Multitask GANs for Semantic Segmentation and Depth Completion With Cycle Consistency", *IEEE Transactions on Neural Networks and Learning Systems*, no. 32(12), pp. 5404- 5415, 2021.

[73] Bai, Y., Zhang, Y., Ding, M. and Ghanem, B.,"Sod-mtgan: Small object detection via multi-task generative adversarial network", *In Proceedings of the European Conference on Computer Vision (ECCV)*, (pp. 206-221), 2018.

[74] Al-Dhabyani, W., Gomaa, M., Khaled, H. and Fahmy, A., "Dataset of breast ultrasound images", *Data in brief*, no. 28, p.104863, 2020.

[75] Thung, K.- H. and Wee, C.- Y., , "A brief review on multi-task learning", *Multimedia Tools and Applications*, no. 77(22), pp. 29705- 29725, 2018.