SYNTHESISING REALISTIC MEDICAL IMAGES USING GENERATIVE ADVERSARIAL NETWORKS

MANOJ GURURAJU

Final Thesis Report

DECEMBER 2020

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Liverpool John Moores University, Liverpool, England and UpGrad, Mumbai, India, for providing opportunities and support throughout the study.

I would like to express my deep and sincere thanks to Dr Ahmed Kaky and Dr Rupal Bhargava for their guidance, motivation, and immense knowledge, which carried me through all the stages of the desertion

I am incredibly grateful to my parents for their love, prayers, caring, and sacrifices to prepare me for the future and motivate me throughout my journey. I thank my whole family and friends for supporting my intentions of taking up masters and understanding my absence in many events.

I would thank my thesis supervisor, Mr Shah Ayub Quadri, for providing valuable feedback and suggestion during the course timeline. I thank all the research advisors and mentors who helped me understand the research domain in the course timeline. I thank the UpGrad team for their support in resolving all issues.

Finally, I would like to extend my gratitude to all of my cohort mates who assisted me in sharing ideas and helpful information during the course.

ABSTRACT

Generative adversarial network (GANs) is breakthrough technology capable of generating realistic-looking images. This technology has revolutionised various domains by its application, such as fashion industries, advertisements industries, to name a few. The adversarial learning approach of the networks can support the medical domain on multiple applications such as image synthesis, image translations, denoising and many more. Among these applications, medical image synthesis can significantly help the healthcare domain by solving the data scarcity problem of medical image datasets. The medical images generated by generative networks can boost the model accuracy and reliability. Past studies proved that GANs could generate realistic medical images, but they were unreliable as they were not anatomically correct. In this research, GANs are applied to generate the liver images. Precisely StyleGAN2 ADA is exercised to generate the images. The feasibility of the network's created synthetic liver images from random noise was assessed using Frechet inception distance and Perceptual path length metrics, which indicate the generated images' visual acuity. Experiment results suggest that StyleGAN2 can generate realistic-looking images and rich in diversity. The best model has achieved pretty good FID and PPL scores. However, visual analysis reveals that not all the generated images were usable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GAN…………....Generative Adversarial Network

DCGAN………..Deep Convolutional Generative Adversarial Network

WGAN…………Wasserstein GAN

BEGAN………..Boundary Equilibrium Generative Adversarial Networks

LAPGAN……....Laplacian Generative Adversarial Network

PGAN………….Progressive Growing of GANs

CatGAN………..Category-aware Generative Adversarial Networks

BiGAN………....Bidirectional GAN

InfoGAN……….Information Maximizing GAN

ACGAN………..Auxiliary Classifier GANs

CGAN………….Conditional generative adversarial network

AUC……………Area under the curve

SWD…………...Sliced Wasserstein Distance

SSIM…………..Structural Similarity Index

PSNR…………. Peak signal-to-noise ratio

VIF…………….Variance inflation factor

UQI…………….Universal Quality Index

FID……………..Frechet Inception Distance

PPL……………..Perceptual path length

CAD……………Computer-aided diagnoses

JPEG……………Joint Photographic Experts Group

PNG…………….Portable Network Graphics

NIfTI……………Neuroimaging Informatics Technology Initiative

GPU…………….Graphical Processing Unit

ADA…………… Adaptive Discriminator Augmentation

CHAPTER 1

INTRODUCTION

1.1 Introduction

The notion of medical imaging was started in 1895 after the invention of the x-ray by Wilhelm Rontgen. The x-ray concept is based on transmitting ionising radiation through the body and projecting it onto a photosensitive plate placed behind it. Due to different densities of tissue within the body would be detected and show if there are any abnormalities. In the 1970s, computer tomography (CT) was developed on the principles of taking multiple images of slices of the body and then putting them together using a computer that helps visualise the body's internal structure. Also, in the 1970s, magnetic resonance imaging (MRI) was developed on the concept of using solid magnetic forces that are used to examine the problems in the tissues of the body. Medical imaging has been improved significantly after the invention of the x-ray device.

Medical imaging is a crucial process for patient diagnosis. Doctors can gauge the patient's condition by analysing the medical images and taking timely actions. Computer-aided diagnoses (CAD) are helping the medical imaging sector by assisting the diagnoses of the medical images. Medical imaging has contributed tremendously to increasing the accuracy of the diagnosis, and because of these technologies, there is less need to do exploratory surgeries. Various medical instruments, such as optical coherence tomography (OCT), ultrasonography (US), positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI), are to mention a few. CAD processes the images obtained by these devices by using computer vision techniques to perform tasks such as segmentation, anomaly detection, classification.

Deep learning techniques employed in CAD would improve the efficiency of time consumed for diagnoses and accuracy, thus helping healthcare institutes cater services efficiently and promptly. Though, the deep learning models' performance heavily relies on the quantity and quality of the dataset. Obtaining medical images for the training of the deep learning models is dearer because these training datasets should be curated from a wide range of individuals, such as different age groups, sex, and health conditions, followed by pre-processing the collected images annotated by the domain experts. Also, acquiring the medical data would impose risk factors such as allergies to dye and exposure to radiation can cause multiple side effects for the individuals.

Advancements in computer vision techniques such as generative adversarial networks (GANs) (Goodfellow et al., 2014) have significantly improved the computer vision domain. GANs have two adversarial networks, generator networks and the discriminator network. These networks would play the minimax game between each other. The generator network maximises specific objective functions while the discriminator minimises the same objective function. This behaviour of the network gave the terminology adversarial.

After GANs gained attention, several studies were carried out to generate the medical image using GANs. (Zhang et al., 2018) trained GANs model that can generate the thyroid tissue images (Baur et al., 2018) successfully developed a model to generate realistic-looking skin lesion images with high resolution and outperformed state-of-the-art model. The purpose of this study is to generate realistic-looking liver images using state-of-the-art architecture StyleGAN2 (Karras et al., 2020) introduced by Nvidia that can generate true to life images and evaluate the generated images using the metric Perceptual Path Length (PPL) and Frechet Inception Distance (FID).

1.2 Problem statement

Applying deep learning techniques into medical image analysis would be a feasible and most effective application that can help the hospital provide efficient service to patients. The primary issue blocking the application of deep learning in medical imaging is the lack of medical images available for training the models. Numerous approaches have been taken to mitigate this issue by applying typical data augmentation like flipping, adding noise, rotation, translation, contrast, saturation, colour augmentation, brightness, scaling, cropping, and other techniques like webly supervised learning to name a few. These techniques can potentially increase the data set, but they cannot add more diversified data. For instance, the ideal medical dataset should contain data from various ethnicities, sizes and shapes, and age groups to make the data more diverse, and datasets are often highly imbalanced due to rare pathologies. A diverse dataset can help train the CAD models to effectively perform as it had been trained in multiple categories the one can anticipate.

Incorporating GANs can solve the scarcity of the data by generating synthetic images using the actual ones. Based on previous research, the GANs have been successfully developed true-to-life images with high resolution using the multiple GAN architecture. However, these images are easily classified as synthetic images from experts due to variation in intensities, anatomically incorrect, presence of artifacts and Et cetera. This study aims to implement StyleGAN2 (Karras et al., 2020) architecture for generating realistic-looking liver images.

1.3 Aims and objective

The research aims to develop a model that can generate realistic-looking medical images that are anatomically correct and evaluate the generated images by using reliable and robust metrics. This research intends to explore the StyleGAN2 architecture and generate liver images.

The following objectives are created based on the study's goal:

- To employ the StyleGAN2 ADA architecture to generate realistic-looking medical images with high resolution and free from artifacts.
- To use limited data samples for understanding the diversity in generated images.
- To evaluate the ability of the model using robust metrics.

1.4 Research questions

The research questions mentioned below are based on a previous literature evaluation in the field of medical images synthesis using GANs.

- Can we generate whole organ images instead of generating regions of interest?
- Is there an information gain in the synthetic samples over the actual training dataset?
- Is gain is higher than using standard data augmentation?
- How many training images are required to obtain reliable generative models?
- Are synthetic images have issues with anatomic accuracy and signal quality?
- Can we develop models for 3D brain volumes?

- Can GANs evaluate based on the tasks rather than generic metrics?
- Is it possible to restore MR images acquired with certain artefacts such as motion, especially in a paediatric setting?
- GANs application as makeup removal, can it be extended to medical imaging with applications in improving bone x-ray images by removal of artefacts such as casts to facilitate enhanced viewing
- Is it possible to generate medical images based on text descriptions?

1.5 Significance of the study

This study contributes to generating a realistic-looking medical image with high resolution that is anatomically correct. Applying deep learning techniques in any field requires a good amount of data with diversified data samples to achieve good performance. However, medical image datasets would generally suffer from fewer samples, highly imbalanced classes, and samples limited by sex, region, and age group. Acquiring these medical data involve many complexities. Collecting ethnically diverse data needs coordination from different research facilities worldwide to gather data.

The acquired images from the candidates are mandatorily subjected to remove the personally identifiable information. (Lotan et al., 2020) work has emphasised the privacy of the medical images, what are the different privacy policies used in the US, European region. As per the research paper, the current privacy policy like HIPAA used by the US is outdated and needs to change the policies accounting for the advancement in AI application in medical imaging. Mismanagement in the handling of these data can breach the patient’s privacy. After removing the identifiable data, it needs to be annotated by a radiologist, aggravating the data collection. The process of obtaining the data might also add risk factors to the health of the individuals, such as exposure to radiation. Using GANs to synthesise medical images can help solve the significant problems mentioned.

1.6 Structure of the study

The remainder of the thesis is structured as follows. Chapter 1, section 1.1 briefly introduced the evolution of medical imaging and its impact on the diagnosis. It was followed by a problem statement discussing various problems gathering the medical image in section 1.2. Section 1.3 explains the aims and objectives of this study, followed by all the research questions gathered from the previous research in section 1.4. The significance of this study is being explained in section 1.5.

Chapter 2 presents detailed information about the previous research in the domain. The research papers have been classified into different sections based on the application of the GANs. Section 1.1 briefly introduce the medical imaging domain. Followed by various applications of the GANs in medical imaging and subsections from 2.2.1 to 2.2.4 highlights research carried out on the respective application category. Section 2.4 presents a discussion of the overall research in the field.

Chapter 3 explains the approach that has been planned for carrying out the research. Section 3.1 briefly introduce the chapter. Section 3.2 captures complete methodologies and is divided into multiple subsections. Subsection 3.2.1 presents the process carried out in data selection, subsection 3.2.2 explains the pre-processing techniques followed, subsection 3.2.3 explains the architecture used for the experiment, and subsection 3.2.4 presents that metric used for evaluating the experiments. Summarisation has been captured in section 3.3

Chapter 4 describes the implementation that's been carried out in this experiment. The chapter introduces and explains the data pre-processing techniques in section 4.2.2, model building in section 4.2.2. The chapter also details the experimentation setup in section 4.2.3, evaluation metrics in section 4.2.4 and summarises the entire section in the end.

Chapter 5 captures the results, limitations and discussions about the experiment. The chapter starts with an introduction in section 5.1, followed by the results and discussion in sections 5.2 and ends with a summary in section 5.4

Chapter 6, titled with conclusions and recommendations, is divided into four subsections. Section 6.1 briefly introduce the chapter, section 6.2 details the conclusion derived from the experiment, section 6.3 captures details of the contribution made to the research field and section 6.4 capture the future recommendations that can be considered.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Medical images play a crucial role in diagnosing patient conditions and treatment. Medical images can be obtained from multiple modalities such as computed tomography (CT), X-rays, magnetic resonance imaging (MRI), mammogram, positron emission tomography (PET), to name a few. Before any significant surgeries on internal organs or understanding the cause of the health issues, Doctors subject the patients to get medical images and use the information for providing the treatment. The advancement of the deep learning domain has been applied in computer-aided diagnosis (CAD). Deep learning can assist radiologists in providing the best medical facilities by evaluating the reports quickly and avoiding misdiagnosis. However, the deep learning system required extensive medical data collection to build accurate and efficient tasks. For instance, if a deep learning model needs to segment the tumours in medical images, the models need to be trained with diverse data with different sizes and shapes of tumours observed in the organs to perform well in critical scenarios.

Most of the available datasets do not have a significant number of images. They are highly imbalanced across the classes, and acquiring these images poses a tremendous challenge to the patient's health, such as allergies to dyes used to get the images and exposure to radiation can cause secondary cancers. Before creating medical image datasets, it is mandatory to get consent from the patient's to use the data and remove all the personal identifications from data to protect the patient's privacy. A dataset must be diverse in terms of having data from different genders, different age groups, different ethnicity and various health conditions, and Such datasets can improve the performance of the deep learning models due to their rich diversity. Annotating these acquired data from radiologists would be costlier and time-consuming, causing further challenges in data collection. We can use deep learning techniques to generate medical images to solve these challenges.

Medical image synthesis is one of the best problem statements that have been under investigation for some time in the deep learning domain. In recent years many researchers have significantly contributed to unsupervised medical image synthesis by using various deep learning technologies which would solve the problems such as insufficient data for training deep learning models, class imbalance, denoising of the medical images from faulty medical scans, image translations of one image modality to other and many more.

Many studies also proved that synthetically generated images are realistic, and radiologists often find it difficult to distinguish between authentic medical images and synthesised images. Numerous research has been carried out in medical image synthesis using Generative adversarial networks (GANs).

2.2 Application of GANs in medical

GANs are comprised of two different networks Generator and Discriminator. There are multiple applications based on these networks. GANs can be applied to generate new images using a small dataset, convert the image from one modality to another, denoise the existing image based on the generator network where the networks the properties of the source image and try to generate similar kind. GANs can also classify the medical images, identify the abnormalities, and judge the medical image's quality based on the discriminator network where the network has already learnt to classify the images as real or fake (Yi et al., 2019). Multiple applications of GANs in medical images are shown in Figure 1.

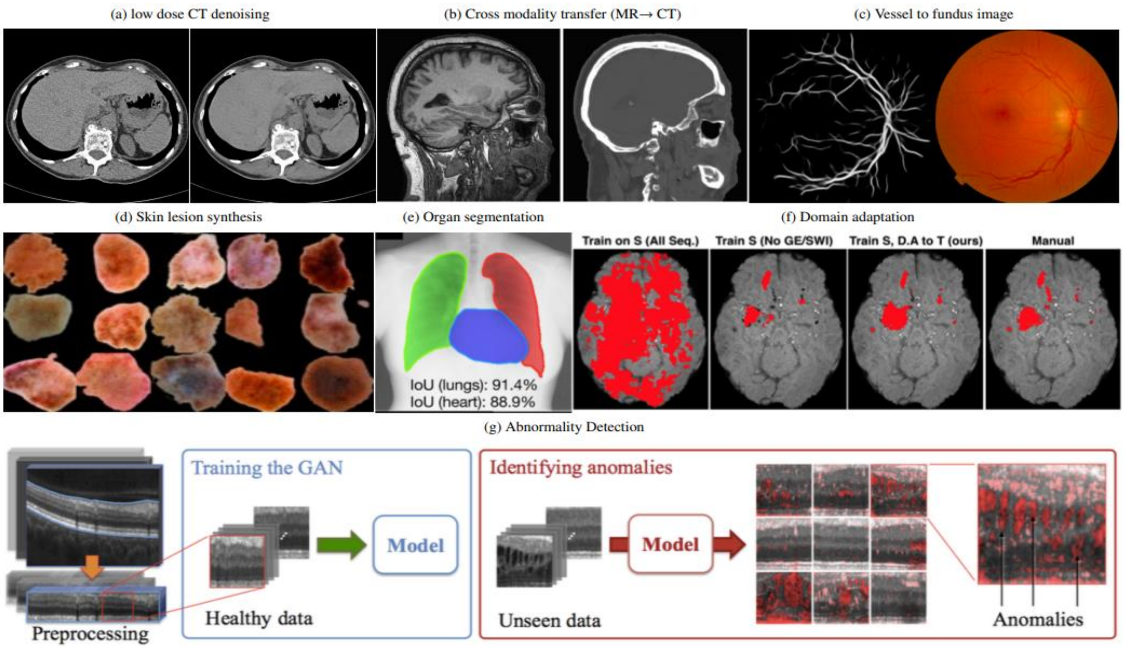


Figure 1: Explaining the applications of the GANs. The figure is cropped from (Yi et al., 2019)

2.2.1 Synthesising medical images

Due to the scarcity of medical images, several approaches are tried to mitigate the problem by using classic image augmentation techniques such as positional and colour augmentation. This augmentation increases the quantity of the dataset but does not increase the quality of the images in terms of providing different sizes, shapes, or health conditions. After the advent of GANs proposed by (Goodfellow et al., 2014), these approaches gained much popularity in generating realistic-looking images. Excellent quality of research has been taken using this approach to solve the data scarcity problem in the medical imaging domain.

(Bermudez et al., 2018) had approached the problem by making models to understand low manifolds of T1 brain MRI data obtained from the BLSA dataset and synthesised realistic and unique brain images that do not correlate with training dataset using GANs incorporating up sampling and down sampling architecture. However, the generated images are readily identifiable as the synthetic image due to un-even intensities, artefacts and anatomically incorrect.

The research carried out by (Zhang et al., 2018) was aimed to improve the image classification model accuracy by increasing the data samples generated from multiple GAN models. In this research, they have used DCGAN, WGAN, and BEGAN to generate thyroid images and found that synthetic images can be used to train the classification model to improve the classification model accuracy. For further improvements, the researchers suggested handling the heterogeneity of the images instead of generating images of whole organs instead of cropped ROI images.

Generating medical images with high resolution was one of the significant challenges previously. (Baur et al., 2018) the research was aimed to generate realistic skin lesion images with high resolution and used PGAN for the task. PGANs are known for generating images from low resolution to high resolution and reducing the training time with the help of this approach. The research was successfully generated realistic images of a skin lesion with high quality, but the generated images were closely matched to the training dataset. Also, the research had concluded with multiple questions for further investigation. Would there be any gain in training the classification models using synthetic images over authentic images? If the gain is higher than the standard augmentation methods? Moreover, how many training images are required to generate realistic images?

(Wu et al., 2020) research has successfully generated urine blood cells using the latest StyleGAN2 architecture (Karras et al., 2020) and created an entire dataset named S2RBC-256.

2.2.2 Converting medical images across modalities

Cross modality images synthesis has a significant impact in the medical imaging field. Different image modalities provide specific information that other modalities do not provide. Patients may need to get multiple imaging modalities to diagnose and cater appropriate treatment in most cases. Acquiring all medical images would expose the patients to harmful rays and increase the probability of developing cancer. The ability of the network that can convert images from one modality such as MRI and convert them to CT would essentially save image acquisition time, cost and resources.

(Nie et al., 2018) have successfully estimated MRI to CT on brain images, MRI to CT on pelvic images and 3T to 7T images of the brain. Their research has introduced several novel approaches such as image gradient difference loss to overcome blurry image generation, introduced long term residual connection to ease training, and successfully estimated target images from the source images.

(Beers et al., 2018) PGANs to transform MRI images of multi-modal glioma to generate retinal fundus and Brian MRI of T1 type transformed to T2 type. The research has also identified latent space to understand the image features better.

2.2.3 Denoising

Denoising the images often reduce the requirement to acquire medical imaging again. Especially in pediatric settings where images might often be blurred out due to slight motion denoising might be helpful. Noise in images is also observed due to radiation dose.

(Bermudez et al., 2018) has developed a denoising autoencoder that takes a noisy image and converts it to clear images.

Several other kinds of research have been done on this application to reduce noise from PET, CT modalities, MR reconstruction and many more.

2.2.4 Discriminator based application

As the discriminator network learns the previous typical images, the network can be further used to detect the abnormalities in the medical images and build a classifier model on top of features of the discriminator network. Also, as the discriminator can identify the image quality, it can judge the quality of newly acquired medical images.

2.3 Related research publication

Prior research emphasises the application of GANs in the medical imaging domain. The survey paper (Yi et al., 2019) has captured a good amount of information on using GANs for various applications in medical imaging and shared challenges that one can anticipate in training the GANs. The paper also highlighted various versions of GANs, significant areas where GANs can be applied in medical imaging, and the areas we can look into in the future in the medical imaging domain using GANs.

(Baur et al., 2018)(Beers et al., 2018) papers have successfully demonstrated that PGAN can be utilised to generate realistic-looking images without losing the quality of the image.

(Skandarani et al., 2021) has explained the different architectures of GANs and the issue they solve. They highlighted various tips and tricks for training the GANs and improving their performance by optimising the hyperparameters and has exhibited favourable results in generating medical images using StyleGAN (Karras et al., 2019)

The improvised version of StyleGAN, StyleGAN2 (Karras et al., 2020), was released in the year 2020 and has significantly improved the performance of the GAN and also resolves the problem of artefacts was faced in the previous version of the architecture. (Wu et al., 2020) has utilised the StyleGAN2 architecture to generate the blood cells found in the urine samples and demonstrate a positive result on generating an entire dataset with more images and rich in diversity.

2.4 Summary

Previous studies suggest that applying the basic GAN application were failed to generate the realistic-looking image, and StyleGAN2 has proven its capability to generate the realistic-looking images without any artefacts. As of December 2021, no previous study attempted to generate liver images using StyleGAN2. This study applies StyleGAN2 to generate realistic-looking liver images and understand the generated images quality.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

A generative adversarial network typically consists of two networks that outperform the other. One tries to generate the image, and another tries to identify whether it is natural or synthetic. Numerous researches successfully demonstrated the capability of generating realistic-looking images using GANs. After analysing all the approaches carried out in the past, Few of the GAN architecture that successfully generated true to life images are PGAN and StyleGAN. PGANs are known for the progressive generation of images. PGANs start their training to generate a tiny image block and systematically add new blocks of convolution layers to both the generator and discriminator model, and this architecture has reduced the training time drastically. The StyleGAN is developed on top of the PGANs architecture. However, the changes are made on the generator part, and the discriminator is implemented almost similar to the PGANs. Our approach in this study is to generate the liver images using the latest state of art StyleGAN2 proposed in 2020 by Karras and the team as the updated version of StyleGAN by fixing a few significant issues.

3.2 Methodology

This section explains the approach taken to generate the photorealistic images using GANs.

3.2.1 Data Selection

The dataset for image generation has been obtained from the Liver tumour segmentation challenge (LiTS). The dataset has 220 CT scans containing the liver image with and without tumours. The liver is the familiar site of primary or secondary tumour development. Due to their heterogeneous and diffusive structure, it would be pretty challenging for generating synthetic images of the organ. CT scans contain slices of images with the internal organ of the region of interest. However, we have to select images only with the liver in the slices for image generation.

3.2.2 Data Pre-preparation

CT scan is used to capture the scan of a section of the body region under investigation, and then the slices of images are combined via computers to represent the whole organ. The CT scans images would not be saved in conventional formats such as JPEG or PNG.

The dataset we selected is in NII or NIfTI format. NII format usually contains 3-dimensional data with a massive volume of data capturing the section of the body. However, the objective of our study is to generate the liver images, so we need to carefully select the images which have liver and are in proper shape and clarity. After selecting the appropriate images, Using the data tools provided by StyleGAN2 PyTorch implementation. Provided helps convert the NII image format to uncompressed PNG files, recommended for training and degrade the training performance if the step is skipped. The data tool expects a path to the source of the image and destination and converts the source data into either an image folder or uncompressed zip archive, easily decoded in the training loop.

3.2.3 Model Implementation

The StyleGAN (Karras et al., 2019) has introduced significant changes to GAN architecture to generate photorealistic images. StyleGAN generates images sequentially, originating from simple resolution to massive scale. StyleGAN transforms each level and verifies the visual features exhibited on that level. The standalone mapping and noise layers are responsible for generating random images. StyleGAN architecture has demonstrated impressive results in generating facial images. The significant characteristics of StyleGAN are converting the latent space vector (z) to W space vector using fully connected mapping layers, W space controlling the generator by shifting the parameters by identifying the mean. This feature helped generate stunning images nearly indistinguishable between real and synthetic. However, the generated images are often associated with artifacts, specifically water droplets, making it easy to identify between real and synthetic images. Figure 2 depicts the examples of water droplets in generated images using StyleGAN.

Figure 2. Water droplet like effects caused by instance normalisation in StyleGAN. The figure is sourced from (Karras et al., 2020)

The second version of this architecture, StyleGAN2 (Karras et al., 2020), targeted to resolve the artfact issue faced in the generation of images. StyleGAN2 has implemented a few changes, such as reconstructing the adaptive instance normalisation. Adaptive instance normalisation in the previous version was used to normalise the feature maps by mean and variance either channel-wise or within the feature map and discards the information, causing GANs to estimate this information using previous information resulting in water droplets. The author of StyleGAN2 separated Gaussian noise and adaptive instance normalisation to overcome this issue, countering the behaviours. In StyleGAN2, the author reconstructed adaptive normalisation as weighted demodulation, avoiding generic normalisation in each level, thus improving training speed by 40%. Figure 3 shows the change in the architecture of adaptive instance normalisation. StyleGAN2 has also introduced lazy regularisation that uses the perceptual path length (PPL) metric in the loss function, applied at specific intervals instead of every iteration. PPL often evaluates the change in the generated images for the slight change in the latent space. Adding it to the loss function has efficiently improved the smooth transition in image generation.

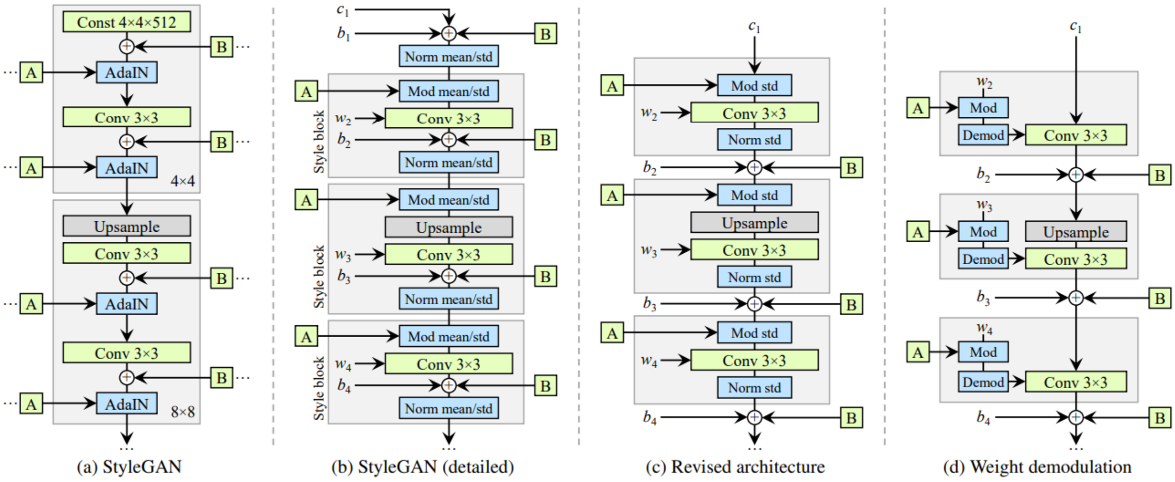


Figure 3. Revised architecture of adaptive instance normalisation in StyleGAN2. The figure is sourced from (Karras et al., 2020).

Another significant change brought in StyleGAN2 is the new generator and discriminator architecture. The StyleGAN2 replace the Progressive GANs architecture with new multi-scale gradient GANs. The PGANs were dealing with many hyperparameter searches in each iteration, creating the bottleneck for the training process. The latest architecture enforces intermediate feature maps to project as images and provide discriminators as additional features. However, StyleGAN2 does not precisely implement the same architecture but uses two different schemes: input/output skip connection and residual net.

In our research, we have taken StyleGAN2 ADA implanted in PyTorch (Karras & Hellsten, n.d.) as a reference to the experimentation. In this study, We are implementing StyleGAN2 architecture for generating realistic-looking images trained from the limited number of input data samples.

3.2.4 Interpretation/Evaluation

Automated GAN metrics such as Frechet Inception Distance (FID), Precession and recall generally use pre-trained classifiers, which are heavily biased by texture and shape detection, identified in the research (Geirhos et al., 2018). FIDs metic is based on the concept of not strictly comparing two images instead, check if the shapes are within the allowed distance. Lower the value of FID is considered a better model. However, (Karras et al., 2020) have demonstrated that images with lower FIDs are not visually correct. The PPL metric can be used for avoiding the issues mentioned above. PPL measure the intensity of change in the generated images for a slight change in the latent space. Lowering the value of PPL is considered a good model. (Karras et al., 2020) have compared the images with FID and PPL. In the comparison, they have identified that PPL can represent the model behaviour more accurately than the FID. We used FID and PPL as metrics to evaluate our study's GAN model.

3.3 Summary

The dataset obtained from the LiTS challenge would be subjected to pre-processing techniques such as converting the image to PNG format and reducing the dimension of the image and necessary rotation. The pre-processed images are then fed to the StyleGAN2 model for training, and the output of the image generations are evaluated using FID and PPL.

CHAPTER 4

IMPLEMENTATION

4.1 Introduction

This chapter explains the data pre-processing steps, model and evaluation metrics implementations performed during the experimentation. As the objective is to generate medical images, there was a requirement to understand the medical image formats and size. Medical images are quite different from the regular image generation tasks and need different approaches to pre-process the data to make it compatible with model implementation. Section 4.2 explains the data pr-processing carried out in this study. Data pre-processing steps were followed by analysing the CT scan slices manually. We need to select the images that have clear images liver carefully. StyleGAN2 was taken from publically available network architecture implemented by the authors of StyleGAN2. The model implementation was done on PyTorch, and the performance of this implantation is approximately 5% to 30% faster than Tensorflow implementation. Evaluation metrics were also sourced from the same code base.

4.2 Model implementation

This section is divided into three sections explaining the data pre-processing techniques, model implementation and evaluation metrics.

4.2.1 data pre-processing

The LiTS dataset utilised for the study consists of 130 CT scans. Most of the medical images uses NII format for storing the CT scan. Each CT scan consists of numerous image slices capturing the data of the section of the body. These CT scans cannot be taken directly for training the GAN models as they are huge and would contain images out of scope for the study. We utilised nib python package for reading the NII format files and converted them to NumPy arrays for better processing.

Fastai has implemented several techniques to pre-process the medical images. We have employed these techniques to pre-process these CT scans. The dataset had high voxel values.

Using the windowing technique, the liver region is highlighted. TensorCTScan provides tools for converting the NII formatted files to JPEG images. The data has been successfully converted to JPEG using TensorCTscan and later to PNG. After converting the data to PNG files, the dataset was carefully selected with clear and appropriate liver structure images. The StyleGAN2 PyTorch implementation recommends using PNG format to decode a training loop easily. Furthermore, the dataset has been converted to the appropriate format using the tools recommended by StyleGAN2.

4.2.2 Model building

Numerous previous researches used various GAN architecture such as DCGAN, WGAN, BEGAN, LAPGAN, PGAN, StlyGAN. They identified that PGANs could generate high-resolution images and StlyeGAN, leveraging PGAN architecture and adding changes. Gaussian noise and adaptive instance normalisation have demonstrated impressive results. Due to adaptive instance normalisation and architecture used in PGANs were generating images with some visual artifacts such as water droplets and phase issues. StyleGAN2 has been developed to mitigate the issues faced by StyleGAN, and StyleGAN, with adaptive discriminator augmentation, generates realistic images even with limited data samples. The authors of StyleGAN2 has made the implementation publically available under restricted licence and can be used for research purposes only. The implementation has been utilised for our research without any changes made to the network.

The StyleGAN2 supports all primary training configurations and performs extensive verification, and the model is 5% to 30% faster than Tensorflow implementation. The implementation would also provide various command-line tools for tweaking the performance.

4.2.3 Experiment details

The StyleGAN2 implementation can be run on both Windows and Linux. The experimentation has been carried out on Google Colab Pro, a cloud GPU platform. The study was carried out using the following tools: Python 3.7, PyTorch 1.7.1 and CUDA 11.2. The model has been trained on Nvidia Tesla P100 for 24 hours and resumed the re-trained the same network for 48 hours.

4.2.4 Evaluation metric

GANs are generally evaluated with automatic evaluation metrics such as FID, Precession and recall, et cetera. However, these metric fails to quantify the visual features accurately, as demonstrated in (Karras et al., 2020) and recommends using PPL. StyleGAN2 ADA provides both FID and PPL for evaluating the quality of images generated using the architecture.

4.3 Summary

This section started with the description of the dataset and models utilised in the study. The original dataset had 130 CT scans and out of which 303 images were extracted by converting the NII format files to JPEG format images. The selected images were then processed using fastai medical imaging tools. These tools effectively highlighted the liver and trimmed it to the 512*512 shape.

Before using these data, they are further fed to the data preparation tool recommended by the StyleGAN2, which converts the images into uncompressed PNG files and captures the metadata such as labels in JSON files. The conversion of images to the dataset by the data preparation tool will help in reducing the training duration. The pre-processed data then catered for training the StyleGAN2 model. Upon training for 72 hours using Tesla P100 GPU, the model achieved the FID score of 52.6 and PPL score of 102.7.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 Introduction

This chapter explains the results obtained from following the methodology mentioned in chapter 3. It explains the quality of generated images using StyleGAN2 using FID and PPL metrics. This chapter also captures the FID score progression during the training, progression of the generated images. It also explains the challenges faced during the implementation and training. Furthermore, the section explains the detailed analysis of the result and what parameters were contributing to the result.

5.2 Results

The proposed model StyleGAN2 ADA has successfully demonstrated its capability in generating realistic images. The dataset used for training the model contained 303 liver images converted to the appropriate format for the training. The model took 72 hours of training using Nvidia Tesla P100 catered by Google Colab Pro. Due to the restriction imposed by the cloud GPU, the execution would be forcefully terminated every 24 hours, the best network from the iterations was re-trained. The best network out of training achieved the FID score of 52.6 and PPL score of 102.7. Lower the FID and PPL scores better the model. Figure 4 depicts the progress of the FID score during the training process.

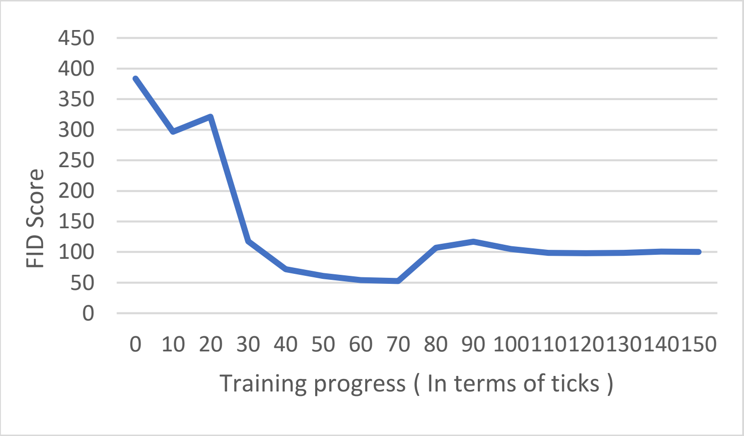


Figure 4. Progress of the FID scores during the training stages expressed in terms of ticks.

Only FID scores were calculated during the training. It is observed that re-training the model from the previous network was not making any significant improvements, and even sometimes, the FID scores were increasing. All parameters in the pre-processing stages are similar, and the only change would be training on the different GPU, which is one of the probable reasons for the increase in FID score during re-training the network (Figure 5).

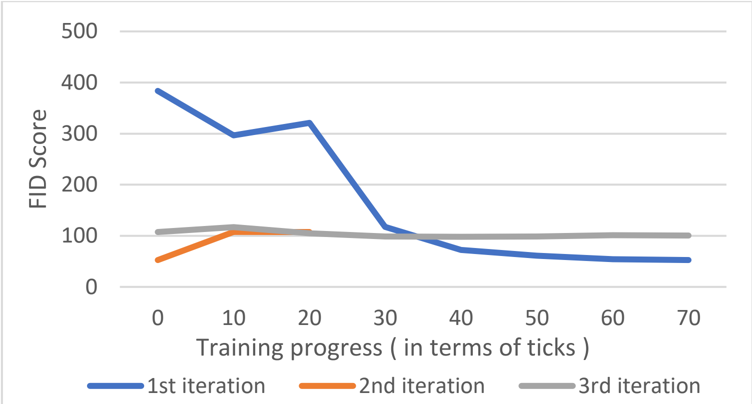
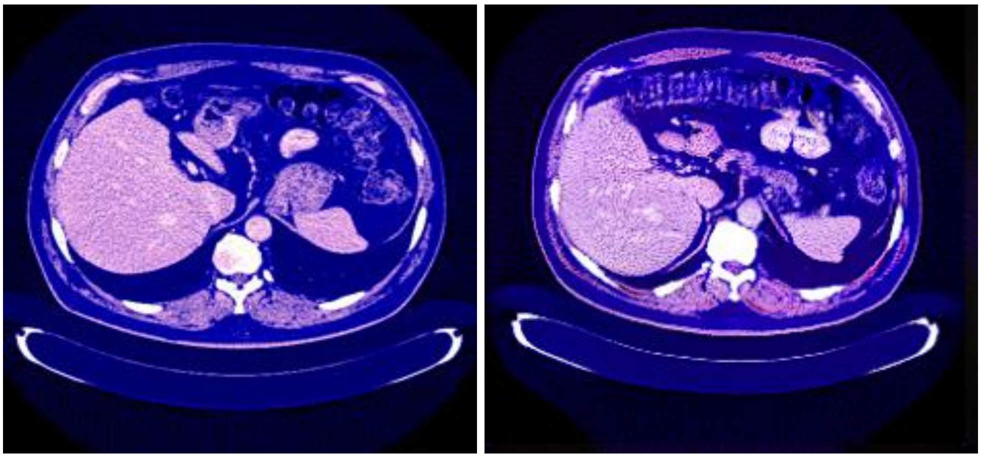


Figure 5. FID scores across iteration of resuming training of the network

The objective of generating realistic-looking medical images with high resolution, free from visual artifacts, has been achieved in this study. After visually inspecting the photos, the network generated a high-resolution, realistic-looking image. In this regard, not all of the images produced are flawless. Figure 6 shows the images generated by GAN and the real image.



(a) Real image (b) StyleGAN2 sample

Figure 6. Real image and Synthetic image samples

Both normal and tumour livers were included in the dataset given to the network for training, and the network was able to capture and output images for both categories. One of the other objectives was to understand the model's behaviour based on the limited data samples. Overall, images generated by the network is satisfactory. However, there are a few cases where the generated images mimic the tumour liver but fail miserably in the process. The dataset used for training consists of only a limited number of images with the liver and caused the generator to learn the exact location of the tumour and the number of tumours and repeat the same behaviour during generation. Figure 7 shows a few examples that mimic the tumour liver with the exact location and number of tumours.

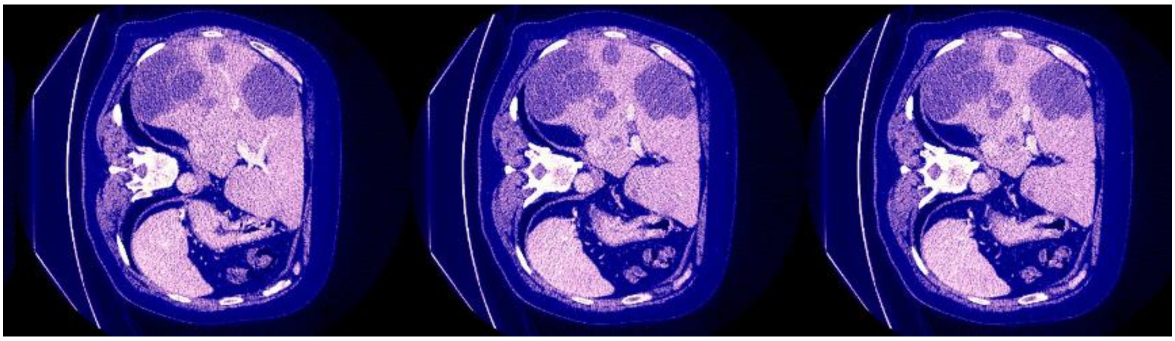
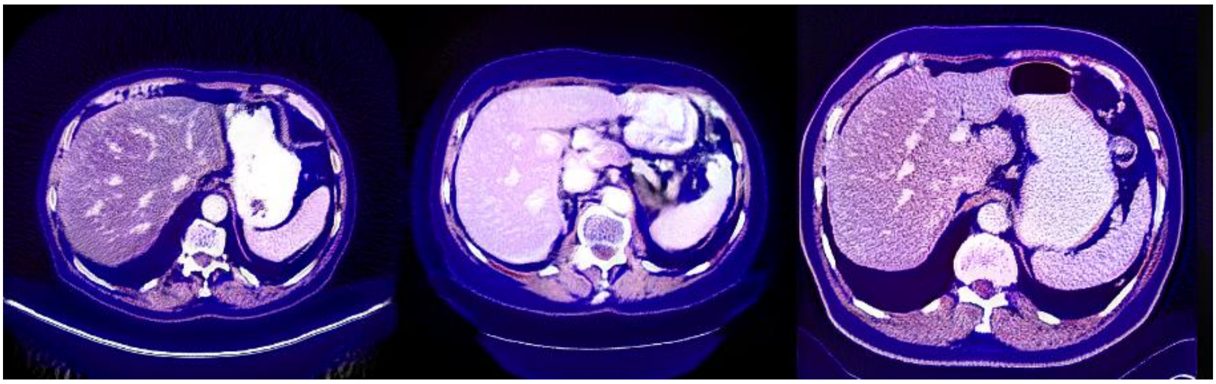
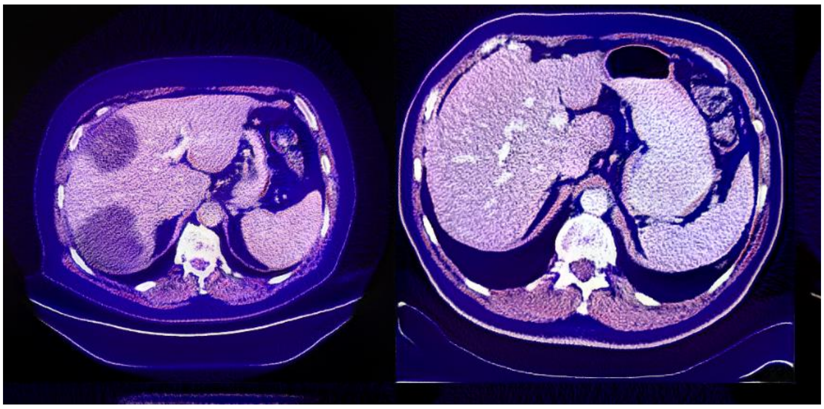


Figure 7. Network generating the images with similar attributes of tumours

Despite the limited number of data samples, the network generated images diverse in size, shape and different classes and attributes of a few classes are pretty similar. Figure 8 demonstrates diverse behaviour with samples.



(a) Different image shapes and sizes.



(b) Liver with tumour (left) and Normal liver (right).

Figure 8. Diverse nature of generated images.

StyleGAN2 uses multi-scale gradient inspired architecture instead of the PGANs, which helps generate images quicker than the previous version StyleGAN and avoids the artifacts in generated images. Figure 9 demonstrates the progression of the images throughout the training.

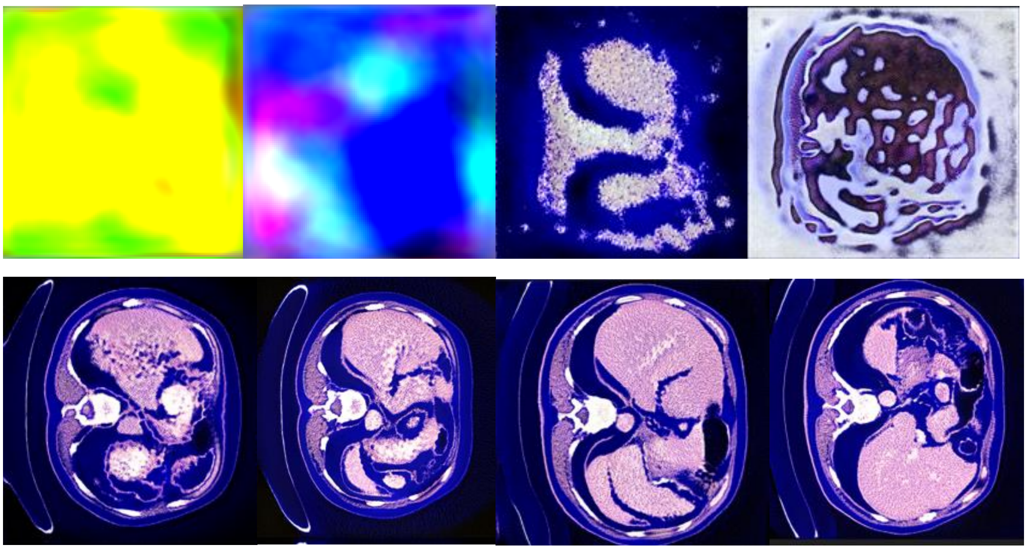


Figure 9. Images generated during the training process.

Evaluating the generated images with metrics such as FID, Precision, and recall does not precisely quantify the image's visual quality. The StyleGAN2 authors recommend using PPL, which depicts the quality of the image close to visual inspection. The best network in this study had achieved the PPL score of 102.7, which implies that the model can generate true to life images.

5.3 Discussions

This research aims to generate realistic looking, high-resolution images using StyleGAN2 ADA architecture with limited data samples and evaluate the quality of the images using appropriate metrics and compare them. The result of the research has successfully fulfilled the objective planned. The generator network is evaluated using FID and PPL, as shown in Table 1. The results achieved during this study imply that the network can generate quality images with further training.

Table 1. Metrics used for evaluation of the model

Model	**FID**	**PPL**
StyleGAN ADA	52.6	102.7

Several challenges were faced throughout the model implementation and training process, which helped identify the factors influencing the training and image generation. Most of the previous researches on medical image synthesis take support from radiologist on the final step for Turing tests. However, if they are involved in the early stages of data-preprocessing, it can help understand the medical imaging and anatomy of the organ, which can further help select the right images for training the GANs. Based on the study, it has been reasoned that input data highly influence the generated images.

GANs networks require extensive computational resources for developing high-quality networks. This research uses Tesla P100 GPU catered from cloud GPU to train the network and generate reasonable images. Though, there were restrictions from the cloud GPU facilitator in execution duration. The execution is limited to 24 hours, and post the network needs to resume the training with the saved snap of the network. Nevertheless, the results obtained after resuming the network were not significant, and the performance was not improved. Instead, it diminished. All the parameters used for resuming the training were the same except the GPU.

Furthermore, this issue can be taken up for an investigation to validate the hypothesis that change in GPU affects training. The research has been limited due to restrictions in the computation resources. If there is a dedicated and capable GPU, the network can be further improved.

5.4 Summary

This chapter started with the result of generating images using StyleGAN2. The StyleGAN2 is a capable architecture for the synthesis of medical images. The training network has achieved the FID score of 52.6 and the PPL score of 102.7. The generated images were not only realistic but also rich in diversity. There were challenges such as domain understanding for selecting the appropriate dataset and extensive computation for developing usable networks that impacted the experimentation.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This section explains the key findings of the research. The section starts with conclusions derived from generating the images using the StyleGAN2 and explains how we can further improve the quality of the generation of these images. In the end, the chapter explains the experiment's limitations and what can be taken forward in future work.

6.2 Discussion and conclusion

This research has successfully generated realistic-looking liver images with the StyleGAN2 model trained only on 303 image samples. The trained network has achieved the FID score of 52.6 and the PPL score of 102.7. Lesser the value of FID and PPL implies better the model. Despite fewer input samples, the trained network generated good quality images and generated diverse data samples. However, The generated images of minority classes such as liver with tumour were reproducing the exact trained images with exact location and number of the tumour. Figure 8 demonstrates a few of the samples generated using the trained network.

Other observations from the experiment are related to the underlying hardware used for training the network. The GANs are generally computationally expensive as two networks are trained to outperform each other. Generally, one network would be trained for the other deep learning tasks to suffice the descent computation. However, as GANs consist of two networks and each network makes lot matrix multiplication during the training process by updating the weights and biases, it requires exceptional GPUs for training the usable network. The GPUs with the performance offered equal to or greater than Nvidia Tesla P100 need to be used for developing the model for StyleGAN2.

6.3 Contribution to the knowledge

Research on a high level is the process of identifying the un-explored approaches in any specific domain and systematically adds results to the vast knowledge base curated over time by many researchers. This research had the objective of using the latest GAN architecture to generate the liver images and evaluate using a robust metric that reasonably represents the generated image's quality. Based on the literature carried out part of the study, no existing research experimented with generating realistic liver images using the StyleGAN2 ADA architecture. This research contributes to understanding the effects of exercising the StyleGAN2 ADA architecture for generating liver images.

6.4 Future recommendation

This study was restricted with a lack of adequate hardware, as seen in the results. As part of this research, Tesla P100 GPUs, which are limited to an execution time of 24 hours and post it all, would be forcefully terminated. However, re-training the same network performance would not be improved. Instead, it has been observed that the performance degrades. Hence, future work can consider training the network on capable hardware without restriction on execution time.

Based on the research results, it is observed that not all generated images were usable. Some images can easily be identified as fake images due to the anatomical structure of the liver. Future work can be considered to evaluate the ratio of usable images over the total generated images to validate the usefulness of the approach to generated medical images.

Another research observation is that the evaluation metrics used did not accurately represent the image generation's quality. Future work can also consider developing a fake medical image detection network that can identify incorrect anatomical structures, artifacts, and images with incorrect intensities. Such a solution could help represent the generator's capability and help improvise the generator's performance.

Due to the COVID-19 impact, the healthcare institutes are flooded with lung CT scan prescriptions. The massive volume of these requests has overloaded healthcare institutes. They estimated the CT scan based on the patient's physical size, age, gender, and other health conditions such as blood oxygen level could be taken for future work.

REFERENCES

Baur, C., Albarqouni, S., & Navab, N. (2018). *Generating Highly Realistic Images of Skin Lesions with GANs*.

Beers, A., Brown, J., Chang, K., Campbell, J. P., Ostmo, S., Chiang, M. F., & Kalpathy-Cramer, J. (2018). High-resolution medical image synthesis using progressively grown generative adversarial networks. *ArXiv Preprint ArXiv:1805.03144*.

Bermudez, C., Plassard, A. J., Davis, L. T., Newton, A. T., Resnick, S. M., & Landman, B. A. (2018). Learning implicit brain MRI manifolds with deep learning. *Medical Imaging 2018: Image Processing*, *10574*, 105741L.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *ArXiv Preprint ArXiv:1811.12231*.

Karras, T., & Hellsten, J. (n.d.). *StyleGAN2 ADA PyTorch*. https://github.com/NVlabs/stylegan2-ada-pytorch

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4401–4410.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analysing and improving the image quality of stylegan. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8110–8119.

Lotan, E., Tschider, C., Sodickson, D. K., Caplan, A. L., Bruno, M., Zhang, B., & Lui, Y. W. (2020). Medical Imaging and Privacy in the Era of Artificial Intelligence: Myth, Fallacy, and the Future. *Journal of the American College of Radiology : JACR*, *17*(9), 1159–1162. https://doi.org/10.1016/j.jacr.2020.04.007

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Medical Image Synthesis with Deep Convolutional Adversarial Networks. *IEEE*

Transactions on Biomedical Engineering, *65*(12), 2720–2730. https://doi.org/10.1109/TBME.2018.2814538

Skandarani, Y., Jodoin, P.-M., & Lalande, A. (2021). GANs for Medical Image Synthesis: An Empirical Study. *ArXiv Preprint ArXiv:2105.05318*.

Wu, X., Wu, Y., Li, X., Zhang, X., & Zhou, X. (2020). *Urine Red Blood Cells Generation Using StyleGAN2 Network*. https://doi.org/10.1145/3433996.3434487

Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, *58*, 101552.

Zhang, Q., Wang, H., Lu, H., Won, D., & Yoon, S. W. (2018). Medical image synthesis with generative adversarial networks for tissue recognition. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 199–207.

APPENDIX A

RESEARCH PROPOSAL

SYNTHESISING REALISTIC MEDICAL IMAGES USING GENERATIVE ADVERSARIAL NETWORKS

MANOJ GURURAJU

Research Proposal

AUGUST 2020

Abstract

Generative adversarial network (GANs) is a breakthrough technology that has capable of generating realistic-looking images. This technology has revolutionised various domains by its application, such as fashion industries, advertisements industries, to name a few. The adversarial learning approach of the networks can support the medical domain on multiple applications such as image synthesis, image translations, denoising and many more. Among these applications, medical image synthesis can significantly help the healthcare domain by solving the data scarcity problem of medical image datasets. The medical images generated by generative networks can boost the model accuracy and reliability. Past studies proved that GANs could generate realistic medical images, but they were unreliable as they were not anatomically correct. In this research, GANs are applied to generate the liver images, Precisely Progressive Growing of GANs (PGANs) and U-net architecture with skip connections are used to generate the images and improve the model by addressing the past research gap. Generated synthetic images of the liver from random noise by the generative network would be evaluated for their practicality by using generic metrics available for GANs followed by the segmentation of the images.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GAN…………....Generative Adversarial Network

DCGAN………..Deep Convolutional Generative Adversarial Network

WGAN…………Wasserstein GAN

BEGAN………..Boundary Equilibrium Generative Adversarial Networks

LAPGAN……....Laplacian Generative Adversarial Network

PGAN………….Progressive Growing of GANs

CatGAN………..Category-aware Generative Adversarial Networks

BiGAN………....Bidirectional GAN

InfoGAN……….Information Maximizing GAN

ACGAN………..Auxiliary Classifier GANs

CGAN………….Conditional generative adversarial network

AUC……………Area under the curve

SWD…………...Sliced Wasserstein Distance

SSIM…………..Structural Similarity Index

PSNR…………. Peak signal-to-noise ratio

VIF…………….Variance inflation factor

UQI…………….Universal Quality Index

LPIPS…………..Learned Perceptual Image Patch Similarity

FID……………..Frechet Inception Distance

CAD……………Computer-aided diagnoses

1. Background:

Medical imaging is a crucial process for patient diagnosis. Doctors can gauge the patient's condition by analysing the medical images and take timely actions. Computer-aided diagnoses (CAD) are helping the medical imaging sector by assisting the diagnoses of the medical images. Various medical instruments, such as optical coherence tomography (OCT), ultrasonography (US), positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI), are to mention a few. CAD processes the images obtained by these devices by using computer vision techniques to perform tasks such as segmentation, anomaly detection, classification.

Deep learning techniques employed in CAD would improve the efficiency of time consumed for diagnoses and accuracy, thus helping healthcare institutes cater services efficiently and promptly. Though, the deep learning models' performance heavily relies on the quantity and quality of the dataset. Obtaining medical images for the training of the deep learning models is dearer because these training datasets should be curated from a wide range of individuals, such as different age groups, sex, and health conditions, followed by pre-processing the collected images annotating by the domain experts. Also, acquiring the medical data would impose risk factors such as allergies to dye and exposure to radiation can cause multiple side effects for the individuals.

Advancements in computer vision techniques such as generative adversarial networks (GANs) (Goodfellow et al., 2014) have significantly improved the computer vision domain. GANs have two adversarial networks, generator networks and the discriminator network. These networks would play the minimax game between each other. The generator network would try to maximise specific objective function while the discriminator minimises the same objective function. This behaviour of the network gave the terminology' adversarial.'

After GANs gaining attention, several studies had carried out to generate the medical image using GANs. (Zhang et al., 2018) trained GANs model that can generate the thyroid tissue images (Baur et al., 2018) successfully developed a model to generated realistic-looking skin lesion images with high resolution and outperformed state-of-the-art model. This research uses various GANs architecture such as PGAN and GANs with U-net architecture to generate medical images with high resolution and evaluate the generated images using the metrics LPIP, SSIM and image segmentation.

2. Related Work:

Medical image synthesis is one of the best problem statements that are under investigation for some time and can assist radiologists in providing the best medical facilities by evaluating the reports quickly and avoiding misdiagnosis. In recent years many researchers have significantly contributed to unsupervised medical image synthesis by using various deep learning technologies which would solve the problems such as insufficient data for training deep learning models, class imbalance, denoising of the medical images from faulty medical scans, image translations of one image modality to other and many more.

Many studies also proved that synthetically generated images are realistic, and radiologists often find it difficult to distinguish between authentic medical images and synthesised images. Much research is happening in applying GANs in medical imaging, and this study has explored few studies and listed in table related research approaches and summary.

Table 1: Related research approaches and summary

Sl. No.	Details
1.	**Year: 2018** **Problem:** Synthesis of thyroid images for boosting thyroid tissue classification. (Zhang et al., 2018) **Purpose:** Generate thyroid tissue images to improve the performance of the classification model **Dataset:** Thyroid OCT images

	Algorithms: DCGAN, WGAN, BEGAN **Evaluation:** Sensitivity, Specificity, Accuracy, AUC **Summary:** Combining both synthesised and original images have improved the classification models performance. However, there is some limitation in this approach due to the heterogeneity of non-thyroid images.
2.	**Year:** 2018 **Problem:** Synthesis of realistic Skin lesion images using GANs. (Baur et al., 2018) **Purpose:** Generating Skin lesion images with high-resolution for resolving class imbalances and improve classification model progress. **Dataset:** 10,000 dermoscopic images of benign and malignant skin lesions included in the ISIC2018 dataset. **Algorithms:** DCGAN, LAPGAN, PGAN **Evaluation:** Sliced Wasserstein Distance (SWD), Visual testing by experts **Summary:** Images generated by PGAN had outperformed the state of the art model by generating high-resolution images of skin lesion
3.	**Year:** 2018 **Problem:** Understanding manifolds of normal brain and generated new images with high quality. (Bermudez et al., 2018) **Purpose:** Generating T1 Brain MRI and denoise the generated MRI and understand the brain features.

	Dataset: Baltimore Longitudinal Study of Aging (BLSA) containing 528 T1 weighted brains MRI. **Algorithms:** GANs with skip connection **Evaluation:** Imaging experts validated the generated images. **Summary:** Vanilla GAN with Skip connection architecture had generated brain images with high quality. The model developed could able to generate highly realistic images but was suffering from anatomical abnormalities
4.	**Year:** 2018 **Problem:** Medical image translation **Purpose:** Developing the framework that can be used for multiple modalities of medical image translation. (Armanious et al., 2020) **Dataset:** PET images, MR images **Algorithms:** CasNet with style-content loss as feature extractor and perceptual loss as the discriminator **Evaluation:**.SSIM, PSNR, VIF, UQI, LPIPS **Summary:** Developed model can translate medical images from MR to CT, Denoise PET and MR motion correction. The highlight of this paper is that they have developed unique architecture and used style content loss for extracting features and perceptual loss of discriminator.
5.	**Year:** 2019 **Problem:** Survey of multiple research activities in medical image synthesis.

	Purpose: Reviewing existing work in this field and provide directions for further research. (Yi et al., 2019) **Dataset:** PET, MR, Retinal images, Skin lesion **Algorithms:** Vanilla GAN, CatGAN, BEGAN, BiGAN, InfoGAN, ACGAN, VAEGAN, CGAN, LAPGAN **Evaluation:** FID, LPIP **Summary:** This study explained the challenges in the training of GANs, Various architectures, variants of generator and architecture, range of applications of GANs in the medical imaging field.
6.	**Year:** 2021 **Problem:** Study on multi-GAN and multi-application study to gauge the benefits of GANs in medical imaging. (Skandarani et al., 2021) **Purpose:** Using GANs to generate cardiac, retina and liver images and understand how similar it is comparing to authentic images **Dataset:** Automated Cardiac Diagnosis Challenge – MRI, Segmentation of the Liver Competition 2007 – CT, Indian Diabetic Retinopathy Image Dataset (IDRiD) **Algorithms:** DCGAN, LSGAN, WGAN, HingeGAN, SPADE GAN, StyleGAN **Evaluation:** FID, Segmentation and comparison with the actual image

	Summary: Research had studied multiple GANs architecture and their capability and also study explained various hacks on the training of GANs and develop a unique evaluation metric to understand if the images generated are reliable

3. Research Questions:

The research questions mentioned below are based on a previous literature evaluation in the field of medical images synthesis using GANs.

- Can we generate whole organ images instead of generating regions of interest?
- Whether there is an information gain in the synthetic samples over the actual training dataset?
- Is gain is higher than using standard data augmentation?
- How many training images are required to obtain reliable generative models?
- Are synthetic images have issues with anatomic accuracy and signal quality?
- Can we develop models for 3D brain volumes?
- Can GANs evaluate based on the tasks rather than generic metrics?
- Is it possible to restore MR images acquired with certain artefacts such as motion, especially in a paediatric setting?
- GANs application as makeup removal, can it be extended to medical imaging with applications in improving bone x-ray images by removal of artefacts such as casts to facilitate enhanced viewing
- Is it possible to generate medical images based on text descriptions?

4. Aim and Objectives:

The research aims to develop a model that can generate realistic looking medical images that are anatomically correct and evaluate the generated images by using reliable and robust metrics. This research intends to explore the latest GANs architecture, employ them to generate medical images, and compare the ability of the models using robust evaluation metrics.

The following objectives are created based on the study’s goal:

- To analyse the latest GANs architecture for generating medical images and implementation of the same.
- To employ the GANs to generate realistic-looking medical images with high resolution and proper anatomical structure.
- To evaluate the ability of the models based on the robust metrics and compare the results.

5. Significance of the Study:

This study is contributing to generating a realistic-looking medical image with high resolution which are anatomically correct. Applying deep learning techniques in any field requires a good amount of data with diversified data samples to achieve good performance. However, medical image datasets would generally suffer from fewer samples, highly imbalanced classes, and samples limited by sex, region, and age group. Procuring medical images for research purposes is a complex process as it includes the challenges of preserving the privacy of an individual by removing all personal identification information from the procured images and also the process of obtaining the data might also add risk factors on the health of the individuals such as exposure to radiations. Using GANs for the synthesis of medical images can help in solving the significant problems mentioned.

6. Expected outcomes:

Research scope will be limited due to time and resource constraints:

- Understanding the data will be limited to knowing the properties such as image size and type of image and will not consider the anatomical study of organ structures and their anomalies.
- The research will be performed on the application level, not on the fundamental level, and research includes the implementation of GANs models capable of generating

medical images. Latest architectures that are computationally expensive are not considered due to lack of resources and time.

- The models will be evaluated based on the generic metrics LPIPS, SSIM and consider segmentation of images as an evaluation metric based on the availability of resources.

7. Research Methodology:

This research will be carried out on the Segmentation of the Liver Competition 2007 (SLIVER07) dataset. The main objective of the research is to synthesise images from the mentioned dataset and evaluate the generated images using appropriate metrics to understand the reliability of the images.

7.1 Overview:

- Comprehend the data set and its properties, apply pre-processing techniques, develop interested GANs models and employ them to generate images.
- Evaluate the generated images using relevant and custom evaluation metrics appropriate for this task and compare the results obtained from various developed models.

7.2 Dataset Description:

Segmentation of the Liver Competition 2007 (SLIVER07) dataset has been used for this research work. The SLIVER07 dataset has 40 liver samples containing at least one tumour, and the rest are pathological. Furthermore, the dataset properties such as dimension need to be understood.

7.2 Pre-processing techniques:

As part of the data exploration, we would obtain helpful information such as image dimensions, demographics, and more critical factors. Based on the analysis, we need to perform the following pre-processing techniques mentioned below.

- Data loading
- Resizing and Normalising the images
- Data augmentation
- Converting the image to the appropriate format

7.3 Modelling Techniques:

As the intent is to generate medical images which are anatomically correct, GANs architecture that has spatial awareness and could able to extract all the features are selected. The architecture selected for the study is listed below.

- PGAN: Progressive growing of GANs (Karras et al., 2017) had made a significant impact in the application of GANs. This architecture had outperformed other GAN architectures such as DCGAN, LAPGAN in the study to synthesise skin lesion images (Baur et al., 2018). One more advantage of using PGAN is that its efficiency in reducing the computational time required for training.
- U-net architecture: A simple U-net architecture with skip connections could help keep the images spatial structure.

7.4 Evaluation metrics

GAN models evaluation should not be grounded only on generic metrics available, But also be evaluated based on the task-specific metric as it does not consider factors specific for the domain. Following metrics are considered for evaluation.

- Learned Perceptual Image Patch Similarity
- Structural Similarity Index
- Segmentation of images.

7.5 Workflow:

1. Loading the dataset
2. Pre-process dataset
3. Model architecture development
4. Training
5. Model evaluation
6. Assess and compare results
7. Conclude research

8. Requirements:

8.1 Software requirements:

Python language will be used to implement the proposal

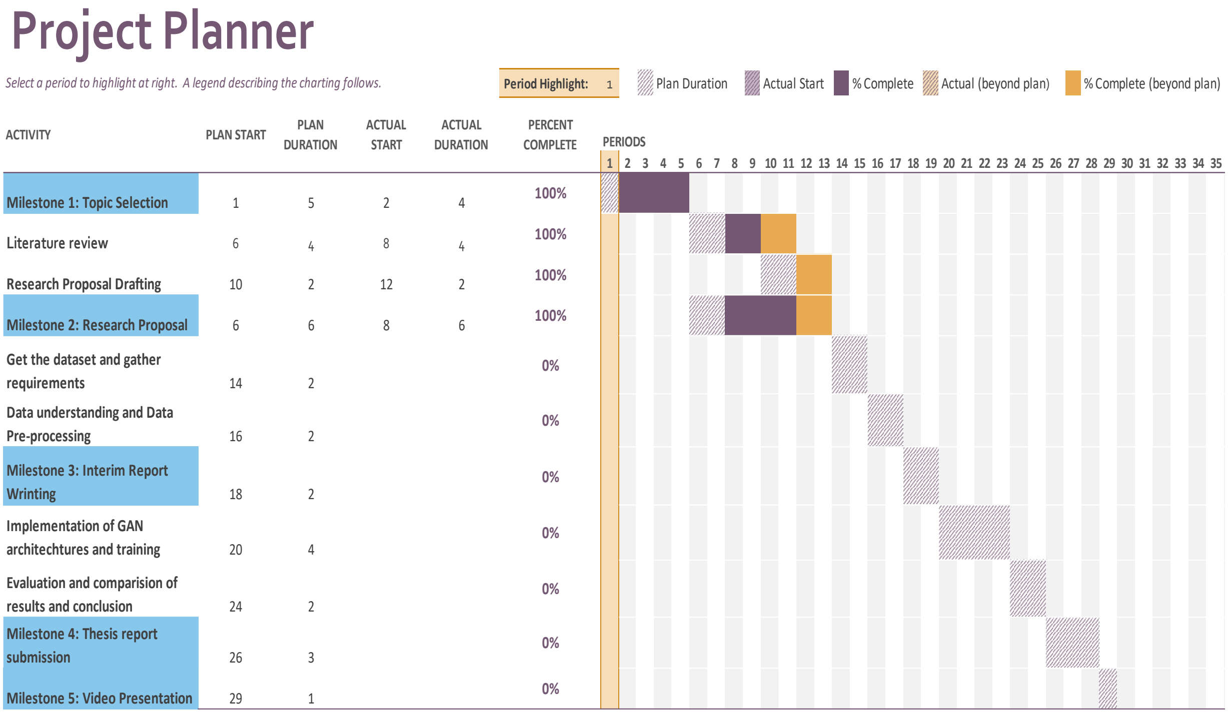
- Python
- Anaconda package manager
- Jupyter notebook
- Python libraries: Numpy, tensorflow, matplotlib, skimage, Keras, imageio, cv2.

8.2 Hardware requirements:

Cloud instance or Laptop with the following configuration:

- 8GB Nvidia GPU
- i7 Processor
- 16 GB RAM

9. Research Plan:


Project Planner

Select a period to highlight at right. A legend describing the charting follows.

ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Milestone 1: Topic Selection	1	5	2	4	100%
Literature review	6	4	8	4	100%
Research Proposal Drafting	10	2	12	2	100%
Milestone 2: Research Proposal	6	6	8	6	100%
Get the dataset and gather requirements	14	2			0%
Data understanding and Data Pre-processing	16	2			0%
Milestone 3: Interim Report Wrinting	18	2			0%
Implementation of GAN architechtures and training	20	4			0%
Evaluation and comparision of results and conclusion	24	2			0%
Milestone 4: Thesis report submission	26	3			0%
Milestone 5: Video Presentation	29	1			0%

Figure 1: Research Plan and Timelines

References:

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., & Yang, B. (2020). MedGAN: Medical image translation using GANs. *Computerized Medical Imaging and Graphics*, *79*, 101684.

Baur, C., Albarqouni, S., & Navab, N. (2018). *Generating Highly Realistic Images of Skin Lesions with GANs*.

Bermudez, C., Plassard, A. J., Davis, L. T., Newton, A. T., Resnick, S. M., & Landman, B. A. (2018). Learning implicit brain MRI manifolds with deep learning. *Medical Imaging 2018: Image Processing*, *10574*, 105741L.

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *ArXiv Preprint ArXiv:1710.10196*.

Skandarani, Y., Jodoin, P.-M., & Lalande, A. (2021). GANs for Medical Image Synthesis: An

Empirical Study. *ArXiv Preprint ArXiv:2105.05318*.

Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, *58*, 101552.

Zhang, Q., Wang, H., Lu, H., Won, D., & Yoon, S. W. (2018). Medical image synthesis with generative adversarial networks for tissue recognition. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 199–207.