Fashion Outfit Design Image Synthesis Using Comparative Study of Generative Adversarial Networks

MSc Research Project
Data Analytics

KARAN JAIN
Student ID: x18183786

School of Computing
National College of Ireland

Supervisor: Dr. Muhammad Iqbal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Karan Jain
Student ID:	x18183786
Programme:	Data Analytics
Year:	2019-20
Module:	MSc Research Project
Supervisor:	Dr. Muhammad Iqbal
Submission Due Date:	17/08/2020
Project Title:	Fashion Outfit Design Image Synthesis Using Comparative Study of Generative Adversarial Networks
Word Count:	7318
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

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

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

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

Fashion Outfit Design Image Synthesis Using Comparative Study of Generative Adversarial Networks

Karan Jain
x18183786

Contents

Abstract

Despite of the capabilities and effects of Artificial Intelligence (AI) it is been observed that there are certain industries which are yet to utilize the immense ability of AI. Fashion industry is slowly making use of machine learning which was used to be a less touched domain of analytics. AI and data analytics both together can bring innovations in the fashion world. A part of this fashion industry is very difficult and that is the fashion designing where new ideas and creations need to be brought in through a line of work. The research is helpful for the fashion designers as it decreases their work pressure using AI to develop fashion images. The Generative Adversarial Network (GAN) is a model to generate samples of images, videos, texts etc. which are more naturalistic. This research focuses on comparative examination using two types of GAN viz- Deep Convolutional Neural Network based GAN (DCGAN) and the Capsule Network based GAN (CapsGAN) to develop new and distinctive images of fashion outfits, jewelries, shoes etc. of both men and women. Both Qualitative and Quantitative methods been used to evaluate the generated images. This resulted in better images generated by CapsGAN than DCGAN. DCGAN showed higher Discriminator loss and in contrast to this CapsGAN showed rise in Generator loss. Altogether it can be stated that CapsGAN performance is much better than DCGAN.

Index terms— Generative Adversarial Networks, Deep Convolutional GAN, Capsule Network GAN, Fashion Industry, Deep Learning

1 Introduction

In this 21st century, Fashion Industry is the most glamorous and multi-billionaire business today. Over the years, the industry has grown rapidly and the larger proportion of the world's population is direct or indirect customers of this industry. The fashion designers who make novel creative designs of fashion items with their remarkable styling skills are the real heroes and runners of this fashion industry. Moreover, the fashion items designs industry seeks all of the attention from people around the globe and various industries and so, makes the competition in fashion domain more tough to get more business with new fashion designs for top most brands. However, there are certain limits to fashion designers creative abilities to maintain its style outcome and freshness of designs.

In today's world, together AI and machine learning have explored all its branches and implemented in various domains so, the new novel creative fashion items design generated images would be very helpful and impressive. Collaborative study of machine learning and AI in the various domain has boost up the industry. Comparatively fashion domain has been less explored in the data analytics field. The recent achievements of AI in fashion have presented that the industry is capable of implementing the other domains beyond the automation. Using machine learning for fashion, business has grown by utilizing fashion designs classification, sales forecasting, style recommender systems, style predictions etc. However, the generative side of fashion has been less-explored such as generating new fashion item design image of shirt, trouser, shoes, bags, etc.

Generative modelling is much popular and seeking attention because of its abilities to generate new image samples of data in the unsupervised learning environment. Since the invention of Generative Adversarial Networks (GANs) Goodfellow et al. (2014), it has performed better than previous generative models like Latent Dirichlet Allocation (LDA), Variational Autoencoders (VAE), Hidden Markov model etc. The Generative Adversarial Network (GAN) is a combination of two neural networks: a Generator and the Discriminator network. The Generator tries to create similar or realistic images from the data samples of given distribution to create better quality samples in less time. And the Discriminator, tries to classify between real samples and generated samples with the minimum loss. The balance between generator and discriminator should be achieved to generate good output sample by GAN model. In image processing and computer vision fields, the GANs model is always preferred and have been proved in their previous studies by researchers with minimum loss and time. GANs have been proven efficient for generating data samples such as videos, texts, dialogues, images, documents etc. by learning about the samples distribution of target data. Liu et al. (2016)

Plenty of work has been conducted in various domains using implementation of basic GAN models and some advanced versions of different GAN models. A detailed description of study on implementations of various GANs is presented in section (2). The aim of this research is to accomplish a comparative study about two advanced altered versions of GANs; Deep Convolutional Generative Adversarial Network (DCGAN) which is proposed by Li et al. (2018) and the other one is Capsule Networks based Generative Adversarial Network (CapsGAN) which is proposed by Wang et al. (2019a). A Convolutional Neural Network (CNN) is used in the Discriminator model by DCGANs as an alternative to classic multi-layer perceptron to increase the performance of Discriminator. CNNs are always preferred to extract multiple features from an image as it utilizes different convolutional fully connected layers. Contrary to this, the CapsGAN utilize a Capsule Network as a discriminator Choi et al. (2019) and the generator is Deconvolutional Neural Network. For this research, the classic Fashion – MNIST dataset by Xiao et al. (2017) have been finalized after precisely comparing some of the fashion items design image datasets which is in detailed explanation in section 2 . The dataset acquired for this research contains 70,000 greyscale images of 28x28 pixels consisting of 10 classes (Ankle Boot, Bag, Coat, T-shirt/Top, Dress, Shirt, Trousers, Sandal, Pullover, Sneaker) thus it makes suitable for our study. The current literature study of the GANs shows the gap where in the fashion domain on a greyscale image dataset the performances of CapsGAN and DCGAN have not been evaluated. Hence, this study aims to fill in this gap and gives a path for future research in this domain accounting our limitations.

The generated sample images by both CapsGAN and DCGAN have been evaluated qualitatively and quantitively. The images quality have been assessed at different epochs by comparing the images evaluating which advanced version of GAN have performed better. And, to find out GANs quantitively, the loss of Discriminator and Generator is compared at several epochs Wang et al. (2019a) These evaluations of GANs have been discussed in detail in section 5.

1.1 Research Question

The Research Question is :
How effectively can Generative Adversarial Networks enhance the generation of fashion outfit images?
The Research Objectives are :
1 - Can GAN models be used to generate new fashion outfit images which can be alternative to fashion designers ?
2 – Is DCGAN or CapsGAN model performs better on the grey-scale dataset of Fashion domain ?

1.2 Expected Contribution

The main objective of this research is to provide an optimal solution to the fashion industry with the help of this generated images of fashion items design images which could serve as base designs for the runners of this industry like fashion designers. This novel creative fashion items design images would be inspiration for using them in industry which are generated images from GANs through this study. Moreover, the comparative study of DCGAN and CapsGAN for generating new dataset of images of fashion items designs been evaluated on quality and quantity metrics on a greyscale image dataset in fashion domain. By implementing this research in real-time for business usage would help growing the fashion industry with the collaboration of business and technology like machine learning and AI.

1.3 Outlining the structure of report :

This is structure of report and following sections consists of section 2 as Related Work followed by section 3 as Research Methodology, section 4 as Design Specification. section 5 as Implementation, section 6 as Evaluation, section 7 as Conclusion and Future Work and lastly section 8 as Acknowledgements.

2 Related Work

2.1 Research based on Utilizing Analytics in Fashion Domain

Analytics in the fashion domain, was less explored in the past years but now researches are holding grip into the trending market and entering into the competitive world in its own ways. And this is succeed because of advancements in the field of AI, Data Analytics and Machine Learning. Most of the study proposed in the fashion domain is either classification of fashion items and its types or predictions of several features like colour, designs, prints, style, sleeves, etc. Few excellent works are discussed in detail below.

In a research carried out by Chan et al. (2018), the authors through this research brought lights into the huge fashion domain by seeking world's attention and even into the various categories of fashion like clothing attributes, several fashion parameters such as style, body measurements, colour, designs. The author have shown directions for the use of this data for forecasting fashion trends in market, exploring the behaviours and preferences of customer, classification of various items and types, etc. In this study, author have provided motivation for further studies in the fashion market. In the similar research proposed by Giri et al. (2019), the authors have suggested a new Panel Data Particle Filter (PDPF) technique to enhance the model for predicting fashion sales. The study have challenged and proven statistically how well the PDPF method have performed better than the existing ARIMA, SARIMA time series forecasting models. A new and similar approach performed by Silva et al. (2019), for forecasting fashion styles. This study have used a Deep Neural Network (DNN) to build their model which can predict the future fashion trends and can forecast which fashion designs or outfits would remain more demanding in market. The model have used several attributes for prediction and acquired accuracy of 90%. On the contrary side of prediction, a research was conducted on classification model by Eshwar et al. (2016), for classification of the fashion apparels by disintegrating study into different stages like upper body signs and then few style attributes of the outfit. This study have altered the utilization of Random Forest for implementing transfer learning have outperformed SVM model by difference of 6.31% accuracy. A similar analysis on fashion outfit image classification model was experimented by Ma et al. (2020), using CNN on the classic ImageNet dataset. They have used Google Net Framework to pre trained the dataset for reducing the training time of the model. After continuous experiments, the average accuracy they achieved was 62%. A recommendation model for fashion designs was proposed by Barsim et al. (2018) presenting a new hybrid method with the help of Association Rule in data mining. The authors have proposed a hybrid model that combines fashion-brand and brand-feature process. The evaluation metrics of the hybrid model outperforms the best F measure score when compared to the existing ones.

Now Fashion Analytics slowly hold the grip into Generative Modelling. In an excellent work conducted by Yu et al. (2019), this research have utilize Generative Adversarial Networks (GANs) wherein the input image was plain fashion outfit and a style to generate a new fashion design image. However, there are few drawbacks to this study like generated images resolution was not good, the real colour transfer from the style image onto the clothing was complex. Furthermore, a novel approach was proposed by Zhai and Zhai (2018) to re-dress the fashion model in a given image. The input image was an image of model wearing outfit and a sentence description. It was named as FashionGAN by the authors. This work uses 2 GAN models and ones output acts as input to second GAN model. The 2nd GAN models generated output is a perfect design transferred image. However, the model have few drawbacks like the GAN model fails to preserve the background of fashion model wearing clothes and the study was more focussed on obtaining the shape of the model. A similar analysis was conducted by Han et al. (2017), here the main focus was to generate fashion apparel scores mainly based on combination of outfit like dress goes with the shoes. This study have used CNN for feature encoding and attain an accuracy of 77%. Another fascinating study carried out by Vasileva et al. (2018), the dataset used for this work was consisting of images of outfits and accessories. The GAN model generates an image of perfect look with an outfit and accessories after model mixing and matching the dataset attributes. This study have used LDA model for generating mixed image. All of these researches in the generative domain have shown further directions for immense research and new innovations in the fashion world.

2.2 Research based on Fashion Datasets

In order to conduct research in the field of fashion, the most important part of study is about suitable datasets of fashion items which are available for public use. The study of Han et al. (2017) the author have used a dataset which was collected from fashion website of US. The dataset contains images around 21K outfit. For our research, the dataset has some drawbacks like the images of outfit are captured in different light settings which makes dataset not suitable. And even there were some duplicate images and for deep learning models training total number of images are comparatively less. In the similar context, author Silva et al. (2019) used a dataset of products which are sold by Amazon in the recent years and it contains 80K number of fashion products images. Unfortunately, this dataset is not suitable for this research, as it contains several fashion accessories images and that is not relevant to this study. The another research by author Giri et al. (2019) have used a dataset named Fashion 144K famous for fashion domain. And it contains images of models wearing different apparels in various backgrounds like in stores, houses, streets, random outdoor places etc. In this images, the extra features acts like obstacles in the training process of the GAN model and makes the training more tedious. Hence, using this dataset for this research is not ideal. Similar research in fashion domain have been carried out by Liu et al. (2016) introducing a new dataset of 80K images captured in various locations with 1050 attributes of annotations. However, it contains images of only men's and women's fashion apparels and does not consists of images of shoes, boots and bags etc fashion products which limits the use of this dataset in out research. Also, the images background settings are quite disturbing. Furthermore, a breakthrough dataset of fashion items design presented by author Xiao et al. (2017), consists of 70K greyscale images of 28x28 pixels with annotations of 10 classes. For the deep learning models, this dataset proven to be benchmark dataset. The images in grey tone are focused on just the product images instead of images of models wearing fashion products in different backgrounds. Hence, here the focus is only on images of fashion item designs and it also contains 10 categories of image samples.

Our research is widely focused on dataset having only designs of the fashion item images with plain background instead of the images of models wearing fashion apparels in disturbing background. And the dataset requires images of various fashion items categories like T-shirts, Trousers, Pullovers, Dresses, Bags, Coats, Shoes etc. having more number of categories making research more relevant and enhance performance of GAN models. Thus, considering dataset presented by author (Citation), using this makes suitable for our research.

2.3 Research based on GAN models

Generative Adversarial Networks is a theory which had been introduced by Goodfellow et al. (2014) that implements neural networks through deep learning techniques in order to estimate the distribution of the data which are newly provided as input. The main purpose of GANs is training two networks at the same time – one is Generator and another Discriminator. Generator helps in drawing a noise to the distribution of data which produces sample of data. Discriminator helps in identifying whether the sample is originally generated or generated falsely. On contrary to this Arjovsky and Bottou (2017) states that GANs are not at all stable and does not function properly. As being told that due to the functional shape of a trained discriminator the GAN became unstable. It had been stated that the solution to this problem is to use mode regularizes to maintain discriminators when using a GAN model. From that time onwards many different types of GANs have been used the researchers across the globe in order to boost its function. StackGAN is a concept which had been initiated by Zhang et al. (2017) to generate more naturalistic images of photos. The model uses 2 GANs viz – First taking text as input followed by shaping and colouring the raw images, which then is send as input to second GAN comprising of the texts where high resolution images are produced. In a report by Deshpande et al. (2018) the writer had brought a concept of Wasserstein GAN also known as WGAN in order to hide the problems of GAN instability and mode collapse. Attention-aware GAN was brought into account by Tang et al. (2019) which made use of transfer learning in the discriminator network to focus on a particular part in the input image which needed more specification. Fang et al. (2018) implemented the usage of GAN with Natural Language Processing (NLP) by advancement of Text-to-Text GAN (TT-GAN) which has the ability to generate natural language. Through this research there is more applications of GAN in NLP. In Patel et al. (2017) a research was carried out in medical field which made use of GAN to identify pigmented and non-pigmented skin injuries in an image. In this regard 92% accuracy was obtained. A similar kind of research was carried out for skin injury by Bi et al. (2019) where GAN was implemented and also use of Convolutional-Deconvolutional Neural Network (CDNN) helped in obtaining 78.1% accuracy. Taking into consideration the benefits of Convolutional Neural Network in computer vision area Li et al. (2018) suggested Deep Convolutional Generative Adversarial Neural Networks (DCGAN) which is a Deconvolutional version of GAN for separately trained environment. In order to increase the categorizing power of GAN this model takes into consideration usage of CNN model inside the Discriminator network of the model. The test was carried on three datasets that includes- LSUN dataset, ImageNet dataset, Celebrity faces dataset. In this study it is been found that DCGAN moderates state-of-the-art versions of GAN. A research was carried out by Silva et al. (2019) using DCGAN model to recognize fourteen plant breeds and twenty-six plant infections where a dataset of 54,000 images was provided giving 89.83% accuracy. As the Capsule Network arrived the researches found out how CNN model was moderated by Capsule Networks in classification space. This is because CapsuleGAN model was offered by Wang et al. (2019a) which is a combination of Capsule Nets with GAN. Here, Capsule Nets was applied as Discriminator testing on 2 datasets in semi-supervised environment.

2.4 Research based on GANs Evaluation Metrics

Researchers around the world are trying continuously so hard to propose new altered versions of GAN for enhancing the functionalities of model and should meet the research needs but still there is a challenge in evaluating the GAN models and assessing it. Recently, some evaluation metrics have been proposed theoretically and the progress been achieved but evaluation and implementation of these metrics is monotonous. The evaluation metrics for the output generated images of GANs are both qualitatively evaluation and quantitatively evaluation. In this similar context, the author Xu et al. (2018) has performed excellent work. The author have proposed around 24 quantitative metrics and 5 qualitative metrics. Several parameters for evaluating images quantitatively are Mode Score, Geometry Score, Inception Score, CAFD Score, AM Score etc. Wasserstein GAN is been evaluated by Wasserstein critic score. The author also recommended that a best way to assess and compare the GAN models performance is by comparing the error rates at different epochs. This clearly states that performance of GANs preferred to be evaluated iteration wise. Moreover, the author has enlighten more on the qualitative parameters stating that in the generated images of GAN the visual analysis is most preferred way for assessing quality and the precision. According to the author, the quality check can be compared for the generated images at several epochs and the performance of GANs can be assessed.

2.5 Gaps in Related Work and Conclusion

In this research, we have gone through different versions of GANs that have been proposed in varied studies and implemented are discussed in detail in related work though considering a gap of comparative study of DCGAN and CapsGAN not yet been executed. Moreover, the capacity of generative modelling makes it more challenging and motivates the further research in the Fashion domain. In this study, generating new fashion items design images through greyscale images of Fashion-MNIST dataset and comparative analysis of the GANs is new in this domain and need to be explored. Hence, by this our research objectives needs to be achieved. And even, after thorough literature study the dataset for this research is been finalized.

3 Research Methodology

To carry out this research as a successful project, it is essential to have systematic project plan wherein huge tasks can be executed into smaller plans. The ideal methodology used to conduct any research is Cross-Industry Standard Process for Data Mining (CRISP-DM) Wirth and Hipp (2000). This methodology covers 6 stages namely: Business Understanding, Data Understanding, Data Preparation, Data Modelling, Evaluation and Deployment which covers all the stages for implementing this research project. Following figure 1 represents CRISP-DM methodology.

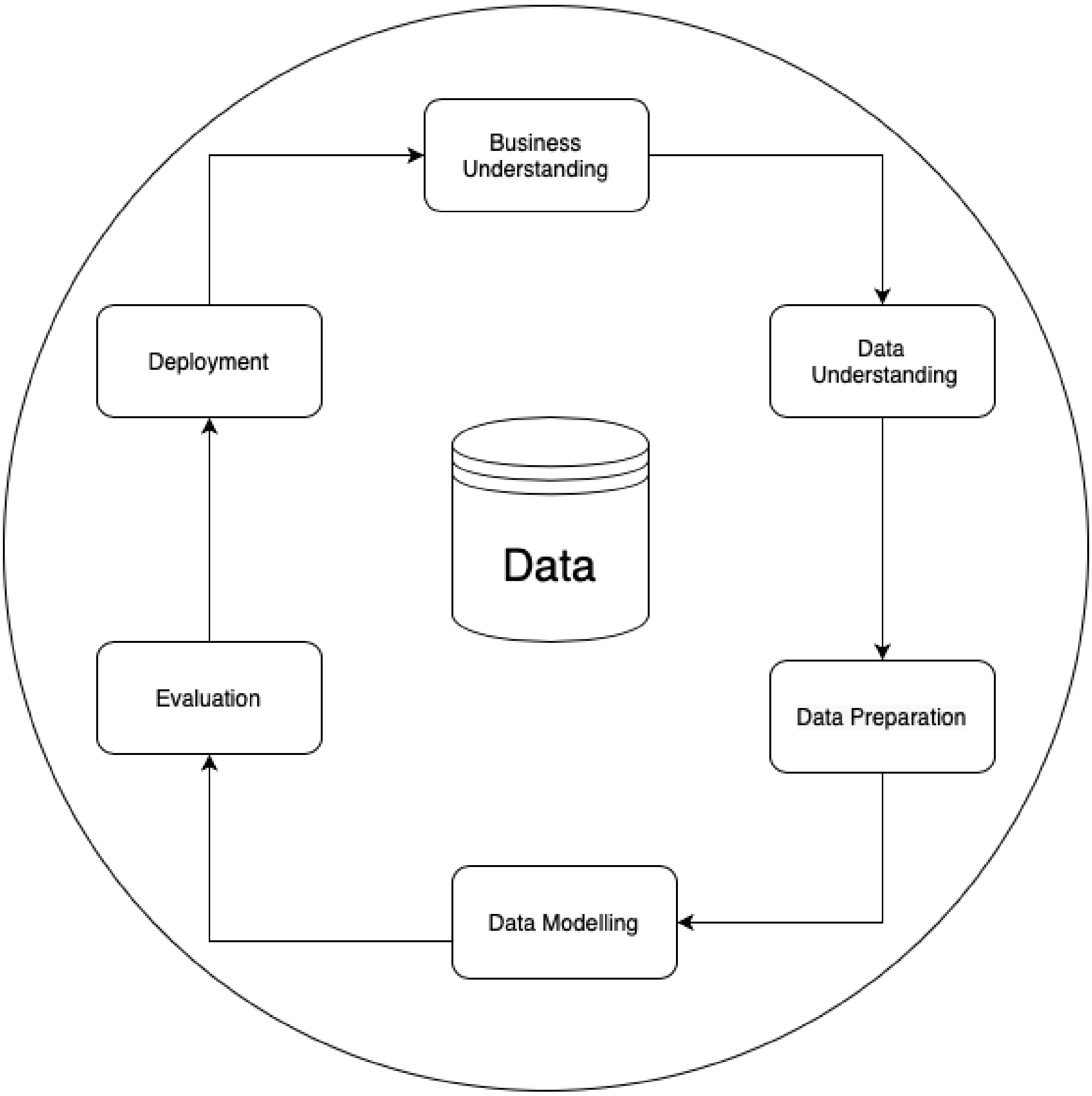

Figure 1: CRISP-DM

[1]

[1]All figures are drawn by own

3.1 Business Understanding

Business understanding is the first stage of the CRISP-DM which is widely focused on understanding the research objectives and finding the business values. This knowledge could be converted into problem statements which can be used to make strategic plans and helps to achieve the objectives of research project. After studying about the fashion industry and analysing the efforts of the fashion designers in creating new fashion designs every time and coping up with the standards of brands in market is tedious job. Hence, this research needs to be achieved with the help of utilizing advancements of data mining, AI and analytics to address this issues in the fashion domain.

Below are the research objectives as stated:

1. Generating various novel images of fashion items design which could help the runners of fashion industry like fashion designers. This generated output images could be named as virtual runners.

2. Motivating the other researchers to explore in similar context to this research and showing directions to work in other domains with the capacity of generative modelling like generating new images through GANs.

3.2 Data Understanding

This stage begins with data acquisition and then understanding the data and finding familiarities within it. For implementing this study, the dataset been utilize is Fashion MNIST classic dataset. The dataset contains 70,000 greyscale images of 28x28 pixels consisting of 10 classes Ankle Boot, Bag, Coat, T-shirt/Top, Dress, Shirt, Trousers, Sandal, Pullover, Sneaker. In this 70,000 images dataset, training images are 60,000 and testing images are 10,000. Another parameter need to be noted while finalizing data is that, the dataset should be available openly for public and as for this study, the data was made available free for open access. Below figure 2 is the snapshot of the dataset consisting of images apparels which are labelled to their respective categories.

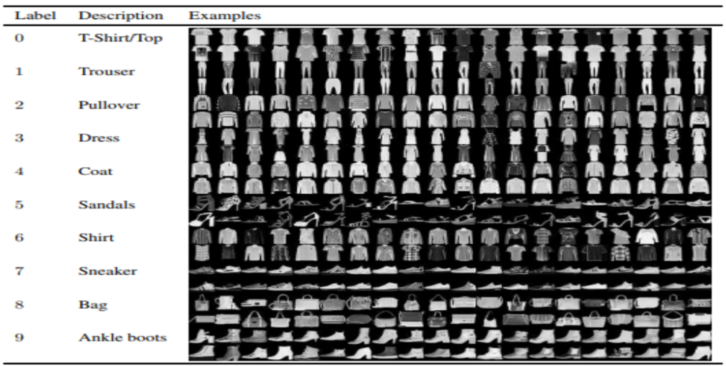

Figure 2: Snapshot of dataset

This benchmarking dataset for machine learning and deep learning researches named Fashion – MNIST contains varied range of categories of images of fashion outfits like clothes, bags, footwear. This quality of dataset makes it ideal and suitable for this research project. The dataset in image pixels looks like: Each row consists of an image in the dataset and the size of images are 28x28 pixels. Each pixels represents the light and dark pixels value of an image wherein high numbers represents darker shades. And this pixel values ranges from 0-255 where 0 means black and 255 means white. The dataset has total 785 columns where 784 columns are from 28x28 pixels consisting each cell one pixel value and there is one column for class label at start of each row. Below figure 3 shows as:

label	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	pixel9	pixel10	pixel11	pixel12	pixel13	pixel14	pixel15	pixel16	pixel17	pixel18
0	0	0	0	0	0	0	0	9	8	0	0	34	29	7	0	11	24	0
1	0	0	0	0	0	0	0	0	0	0	0	209	190	181	150	170	193	180
2	0	0	0	0	0	0	14	53	99	17	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	161	212	138	150	169	164	176	202	255
3	0	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	17
2	0	0	0	0	0	44	105	44	10	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	1	0	0	0	108	25	0	0	0	132	54
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	92	66	0	0	0	0	81	91
3	0	0	0	0	0	0	0	1	0	0	83	142	50	0	0	0	0	85
4	0	0	0	0	0	0	1	1	0	0	21	153	100	88	81	130	50	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
6	0	0	0	0	0	0	1	1	0	0	0	142	122	94	95	136	177	13
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	1	5	0	0	0	91	167	142	143	166	148	0
3	0	0	0	0	0	0	0	0	15	132	121	148	148	145	151	151	154	138
6	0	0	0	0	0	0	0	2	2	2	0	114	255	210	205	253	112	0
4	0	0	0	0	0	0	0	0	0	0	0	0	129	129	113	149	115	0

Figure 3: Dataset in image pixels

3.3 Data Preparation

Data acquisition is the most essential stage in any research project. The dataset should be finalized after performing relevant study in domain as if dataset is not suitable for the project, the desired goal of the research may not be achieved. The EDA and pre-processing of the data is done in Python using TensorFlow. Below are the steps of data been pre-processed for our research project:

3.3.1 Reading the Dataset

Four files are present in the original dataset consists of training images and its labels and testing images along with its labels. The dataset description is as shown in table below. This Fashion – MNIST dataset is already present in TensorFlow as inbuilt dataset which makes the pre-processing step more easy. In TensorFlow, it is more quick to access the dataset directly using this:

File Name	Contents	Examples
train-images-idx3-ubyte.gz	training set images	60,000
train-labels-idx1-ubyte.gz	training set labels	60,000
t10k-images-idx3-ubyte.gz	test set images	10,000
t10k-labels-idx1-ubyte.gz	test set labels	10,000

```
#read the dataset
from tensorflow.examples.tutorials.mnist import input_data
mnist = input_data.read_data_sets("MNIST_Fashion/")
```

3.3.2 Checking on dataset shape

The dataset shape can be defined by shape of an image as each cell consists of an pixel value. So, the total number of rows represents the number of images and total number of pixel values represents number of columns. The shape of training images and testing images are:

```
(60000, 785) (10000, 785)
```

The training images shape (60000,785) shows that there are 60000 sample images which has 785 columns of pixel values.

3.3.3 Analysis of train and test data

After analysing the dataset, the total number of images per categories (0-9) there are 6000 images in training set per category and 1000 images in testing set per category. The below figures 4 and 5 of training and testing data shows the number of images per categories.

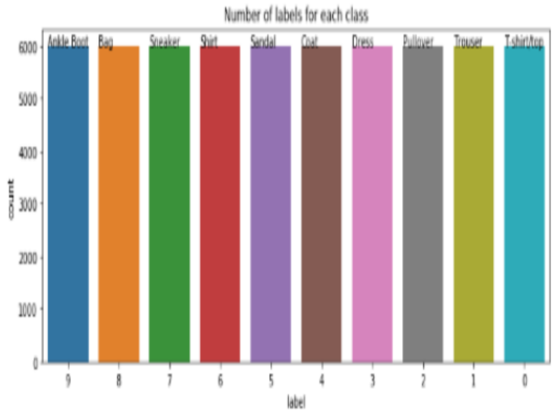

Figure 4: Train data

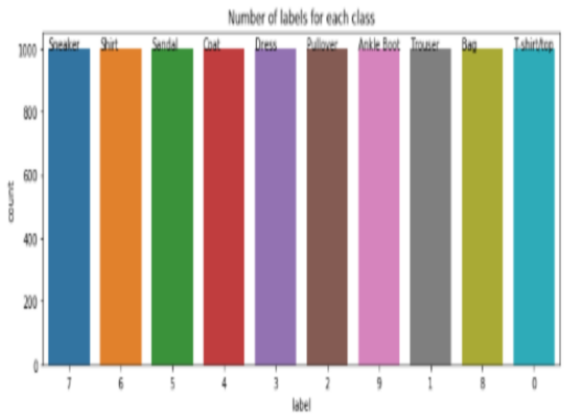

Figure 5: Test data

3.3.4 Null and missing values

In the training data and testing data there were no null values or missing values found. As that may be the reason Fashion – MNIST dataset is preferred as benchmark dataset for researches in deep learning projects.

3.3.5 Normalization of data and Reshaping

The images pixel values lies between a range of 0-255. The input parameters for the deep learning or neural network models must lied between a range of 0-1 for the best performance of model. And this helps in reducing the training time of the model. Hence, the data been normalized to bring all the pixels on a common scale in range of 0 to 1 by dividing every pixel size by 255. Further, after normalization of data it needs to be reshape so that it can be fed as input in the vector form. Moreover, the first layer of DCGAN and CapsGAN takes shape of the input as 28*28*1. Here, 1 represents number of channels as an input. Hence, reshaping of data is performed on both training and testing data so that it can be fed into the model in required form as input. The shape after pre-processing of data of training and testing data are shown below:

```
(60000, 28, 28, 1)        (10000, 28, 28, 1)
```

Hence, after exploring and pre-processing the data now it is ready to be fed the dataset into the model as expected input for the deep learning models.

3.4 Data Modelling

Two altered versions of the GANs are implemented on the dataset after data preparation is done. This 2 advanced versions are Capsule Network based Generative Adversarial Network (CapsGAN) and Deep Convolutional Generative Adversarial Network (DCGAN). The main objective of this research here is to compare which GAN models output and performance is better between CapsGAN and DCGAN on a Fashion – MNIST dataset. And the other objective is to generate new fashion outfits design images using GAN models. For this study, we have compare the results between them at several epochs for analysing and evaluating the model. The detailed implementation of the models is explained in following sections.

3.5 Evaluation

There are varied evaluation parameters for evaluating the generated output samples of this model wherein this research will evaluate based on qualitative and quantitative metrics. For evaluating qualitatively, the generated output sample images from both models will be compared on different epochs. And for quantitative evaluation, the Discriminator loss and the Generator loss will be compared based on outputs generated at several epochs to analyse which model is more suitable for fashion domain.

3.6 Deployment

This research study is been proposed and implemented in academics as part of dissertation. Hence, it is not been deployed in practical for publications. Moreover, this project report can be referred as final deployment.

4 Design Specification

The below figure shows the overall process flow of our research.

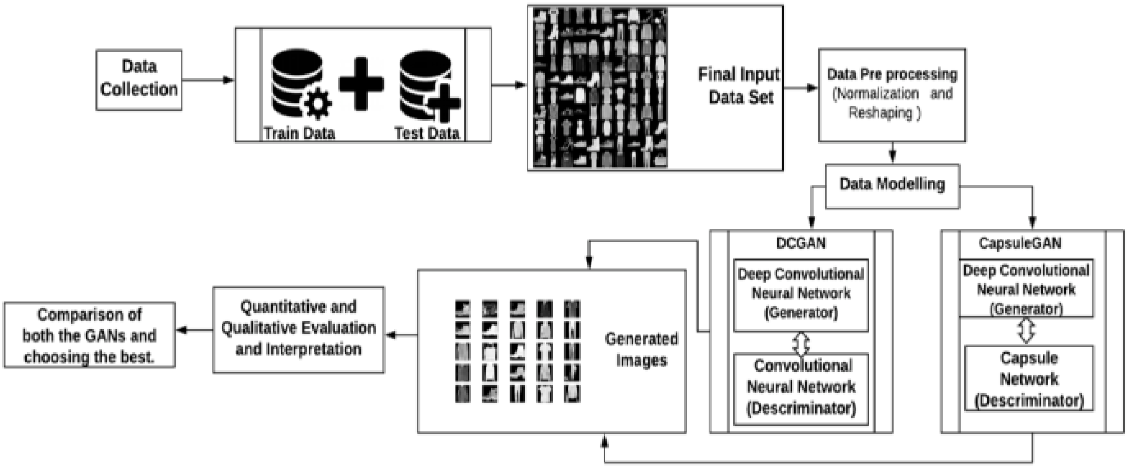

Figure 6: Process Flow Diagram

5 Implementation

The main objective of this research is to generate new fashion outfits images using GAN models and conducting a research on a comparative study of the advanced versions of GANS that is DCGAN and CapsGAN models implemented on the fashion domain. All the implementation of models is carried out in Python using TensorFlow and Keras frameworks. This section focusing on the details of model implementation in our research.

5.1 Generative Adversarial Networks (GANs)

This research implements a model Generative Adversarial Networks (GANs) which is well known for its generative nature. This generative modelling GAN model is basic GAN. The basic GAN comprises of 2 neural networks name as Generator and the Discriminator. The Generator takes the noise as input vector and generates new sample images whereas the job of discriminator is to distinguish between the real image and fake image in a binary output. Both the Generator and Discriminator uses Convolutional Neural Network (CNN). The CNNs used have 2 convolutional layers and pooling layers followed by 2 fully connected layers. In 2nd and 3rd layer, batch normalization have been used and for all layers Leaky ReLU activation function was used followed by last layer of sigmoid function. The hyperparameter K determines the number of epochs applied to the model. The GANs training and implementation is hard as it is said, its simple to recognize a painting than creating new one. The below figure 7 is model summary of GAN.

Both the networks have conv2d that is 2 convolutional layers inside which consisting of 2 - strides of 5x5 filters. The weights of conv2d used 0.02 standard deviation and the bias was

Layer		Output Shape	Parameters
[x(28,28,1), y(10)] -> h_0			
	conv2d_1	14, 14, 11	3036 (5 × 5 × 11 × 11 + 11)
	leaky_relu_1	14, 14, 11	0
	concat_1	14, 14, 21	0
h_0 -> h_1			
	conv2d_2	7, 7, 74	38924 (5 × 5 × 21 × 74 + 74)
	batch_normalization_1	7, 7, 74	148 (74 + 74)
	leaky_relu_2	7, 7, 74	0
	reshape_1	3626	0
	concat_2	3636	0
h_1 -> h_2			
	fully_connected_1	1024	3724288 (3636 × 1024 + 1024)
	batch_normalization_2	1024	2048 (1024 + 1024)
	leaky_relu_3	1024	0
	concat_3	1034	0
h_2 -> h_3			
	fully_connected_2	1	1035 (1034 × 1 + 1)
	sigmoid_1	1	0
Total params: 3,769,479			

Figure 7: Model summary of GAN

kept 0. The slope value of 0.2 was initialized for the Leaky ReLU activation function and the sigmoid function acts as loss function. ADAM optimizer was used for both the networks and for the pooling layers k-size determines the kernel size.

5.2 Deep Convolutional Generative Adversarial Networks (DCGAN)

The DCGAN model is altered advanced version of GAN. Though the researches of GAN is excellent, but when complex datasets applied on model training becomes unstable. The authors Li et al. (2018) have recommended to use CNNs in the discriminator network for stabling the model. The CNN is always preferred for image data in supervised learning but using CNNs in GAN in unsupervised learning would be interesting and exploring new study. The guidelines given by Li et al. (2018) for creating a stable DCGAN model is:
a. Use batch normalization in both networks b. Activation function Leaky ReLU c. Use convolutional layers instead of using pooling layers.

Discriminator: DCGAN uses Convolutional Neural Network (CNN) as a discriminator with convolutional layers and pooling layers. The discriminator network of DCGAN and basic GAN is same. CNN network uses 2 convolutional layers, 2 fully connected layers, batch normalization, Leaky ReLU activation function and the sigmoid function.

Generator: Even the generator have 2 convolutional layers same as discriminator. But this layers are 2D transposed deconvolutional layers. Batch normalization is used in each layer except the last one. ReLU activation function is used in first 3 layers whereas sigmoid function is used in last layer. The model summary of DCGAN is shown in figure 8 below:

Layer		Output Shape	Parameters
[z(100), y(10)] -> h_0			
	fully_connected_1	1024	113664 (110 × 1024 + 1024)
	batch_normalization_1	1024	2048 (1024 + 1024)
	relu_1	1024	0
	concat_1	1034	0
h_0 -> h_1			
	fully_connected_2	6272	6491520 (1034 × 6272 + 6272)
	batch_normalization_2	6272	12544 (6272 + 6272)
	relu_2	6272	0
	reshape_1	7, 7, 128	0
	concat_2	7, 7, 138	0
h_1 -> h_2			
	conv2d_transpose_1	14, 14, 128	441728 (5 × 5 × 128 × 138 + 128)
	batch_normalization_3	14, 14, 128	256 (128 + 128)
	relu_3	14, 14, 128	0
	concat_3	14, 14, 138	0
h_2 -> h_3			
	conv2d_transpose_2	28, 28, 1	3451 (5 × 5 × 1 × 138 + 1)
	sigmoid_1	28, 28, 1	0
Total params: 7,065,211			

Figure 8: Model summary of DCGAN

Similar to basic GAN at some extent, both the networks have conv2d that is 2 convolutional layers inside which consisting of 2 - strides of 5x5 filters. The weights of conv2d used 0.02 standard deviation and the bias was kept 0. For performing batch normalization, the epsilon was 10-5 and the decay factor was 0.9. The slope value of 0.2 was initialized for the Leaky ReLU activation function and the sigmoid function acts as loss function along with logits entropy. 0.0002 ADAM optimizer was initialize for both the networks.

5.3 Capsule Network based Generative Adversarial Network (CapsGAN)

The CapsGAN is also an altered advanced version of basic GAN. The author Wang et al. (2019b), have conducted a study on Capsule Network using as a discriminator in a basic GAN model. The DCGANs architecture is base for CapsGAN as the generators networks in both are same just the discriminator network is different. The ultimate layer that is 16D Capsule net uses dynamic routing in middle of capsules. The discriminator input follows a sequence of stride convolutional layers, capsule network and the output.

Discriminator : Instead of CNN as discriminator network in DCGAN here is Capsule net in CapsGAN. The first convolutional layer uses kernel size 9, 1 stride and 256 filters. It mainly contains 2 Capsule net layers namely Primary-Caps layer and Digit-Caps layer. Along with this it consists of Leaky ReLU activation function, batch normalization, flatten function and Keras Dense Layer. At last, it contains sigmoid function. The detailed layer wise process flow of Capsule net is shown in figure 9 below:

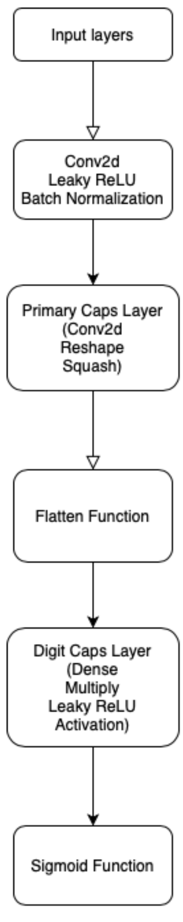

Figure 9: Process flow of Capsule Net

Generator : The generator of CapsGAN and the DCGAN are same using deconvolutional neural network. Hence, the implementation of the generator is in same way. The model summary of CapsGAN is shown in figure 10 below:

Layer (type)	Output Shape	Param #	Connected to
input_4 (InputLayer)	None, 28, 28, 1)	0	
conv1(Conv2D)	(None, 20, 20, 256)	20992	input_4[0][0]
leaky_re_lu_5 (LeakyReLU)	(None, 20, 20, 256)	0	conv1[0][0]
batch_normalization_6 (BatchNormalization)	(None, 20, 20, 256)	1024	leaky_re_lu_5[0][0]
primarycap_conv2 (Conv2D)	(None, 6, 6, 256)	5308672	batch_normalization_6[0][0]
primarycap_reshape (Reshape)	(None, 1152, 8)	0	primarycap_conv2[0][0]
primarycap_squash (Lambda)	(None, 1152, 8)	0	primarycap_reshape[0][0]
batch_normalization_7 (BatchNormalization)	(None, 1152, 8)	32	primarycap_squash[0][0]
flatten_2 (Flatten)	(None, 9216)	0	batch_normalization_7[0][0]
uhat_digitcaps (Dense)	(None, 160)	1474720	flatten_2[0][0]
softmax_digitcaps1 (Activation)	(None, 160)	0	uhat_digitcaps[0][0]
dense_6 (Dense)	(None, 160)	25760	softmax_digitcaps1[0][0]
multiply_4 (Multiply)	(None, 160)	0	uhat_digitcaps[0][0], dense_6[0][0]
leaky_re_lu_6 (LeakyReLU)	(None, 160)	0	multiply_4[0][0]
softmax_digitcaps2 (Activation)	(None, 160)	0	leaky_re_lu_6[0][0]
dense_7(Dense)	(None, 160)	25760	softmax_digitcaps2[0][0]
multiply_5(Multiply)	(None, 160)	0	uhat_digitcaps[0][0] dense_7[0][0]
leaky_re_lu_7(LeakyReLU)	(None, 160)	0	multiply_5[0][0]
softmax_digitcaps3 (Activation)	(None, 160)	0	leaky_re_lu_7[0][0]
dense_8(Dense)	(None, 160)	25760	softmax_digitcaps3[0][0]
multiply_6(Multiply)	(None, 160)	0	uhat_digitcaps[0][0] dense_8[0][0]
leaky_re_lu_8 (LeakyReLU)	(None, 160)	0	multiply_6[0][0]
dense_9 (Dense)	(None, 1)	161	leaky_re_lu_8[0][0]

Total params: 6,882,881
Trainable params: 6,882,353
Non-trainable params: 528

Figure 10: Model summmary of CapsGAN

6 EVALUATION

Evaluation section is the most essential stage for the research through which the implementation can be conclude how well the model is trained and applied. This section evaluates the performance of GAN models. This research objective is mainly focused on a comparative study of DCGAN and CapsGAN, which are two advanced versions of GAN, on the benchmarking Fashion – MNIST dataset. And the other objective is to generate new novel Fashion outfit design images using basic GAN model. For better understanding and interpretation of the results, evaluation is conducted on 2 metrics : qualitative parameter and quantitative parameter. The qualitative evaluation gives visual representation of the output images and quantitative evaluation determines the loss and accuracy of model. The evaluation for all the models is discussed in detail below.

6.1 Qualitative Evaluation

Qualitative evaluation has been conducted on basis of visual analysis of generated images of GANs as suggested by Zhai and Zhai (2018). The quality of images and judging factor is more easy while looking at images with human eyes. Below are the output images of DCGAN and CapsGAN at various epochs carried out to enhance the performance of model.

6.1.1 DCGAN

Below figure 11 are the generated images by DCGAN

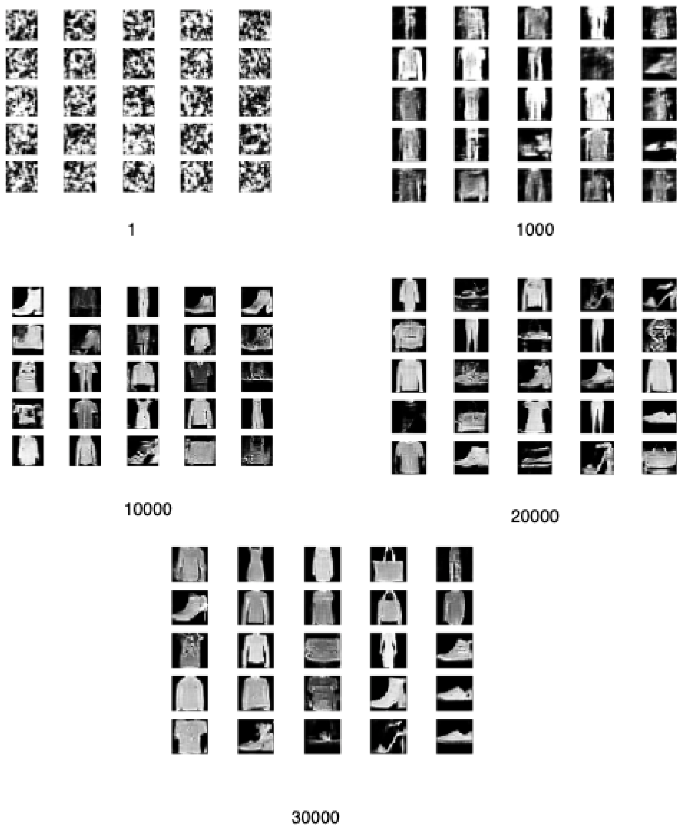

Figure 11: Generated Images by DCGAN

6.1.2 CapsGAN

Below figure 12 are the generated images by CapsGAN

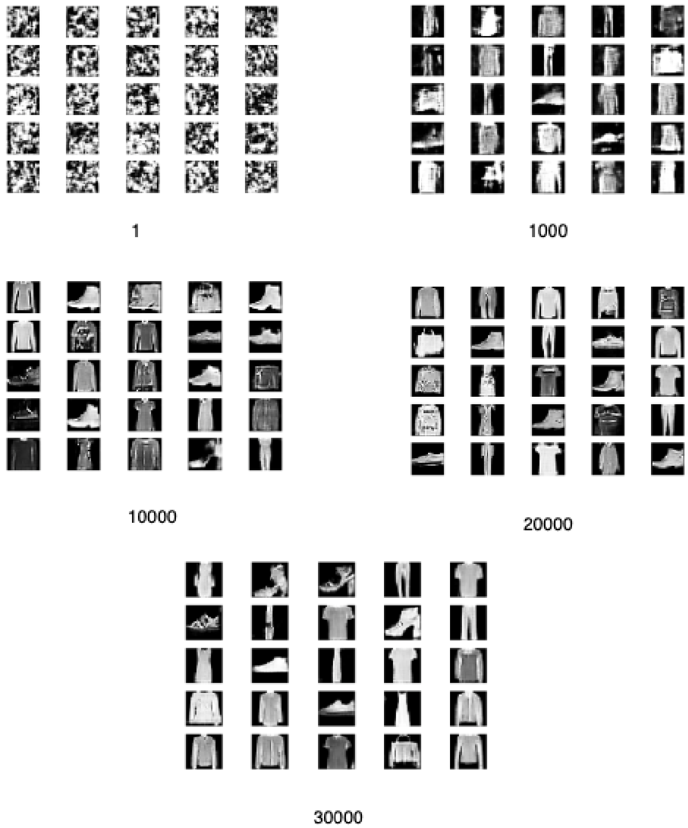

Figure 12: Generated Images by CapsGAN

At the start of training the models, at 1st epoch both images looks similar, while at 1000th epoch, both models have just started holding the grip of shape and designs of fashion images. At 10000th epoch, DCGAN generated images looks more better. While further at 20000th epoch and 30000th epoch, the CapsGAN generated images have better quality and precision of the images compared to DCGAN. Hence, this comparison can be concluded as though DCGAN was quick learner at start and CapsGAN took some time but at the end CapsGAN have totally outperformed the DCGAN generated images quality.

6.2 Quantitative Evaluation

Quantitative evaluation has been conducted on basis of calculating the loss generated by the model at the time of implementing a model. It also determines how well the generator and discriminator have performed and hence, the result can be concluded. The losses of both the models have been compared through graph after 30000th epochs completed for finding out the best model in fashion domain. Below graph shows the loss for both the models.

6.2.1 DCGAN

The below figure 13 is a graph of Loss of DCGAN model

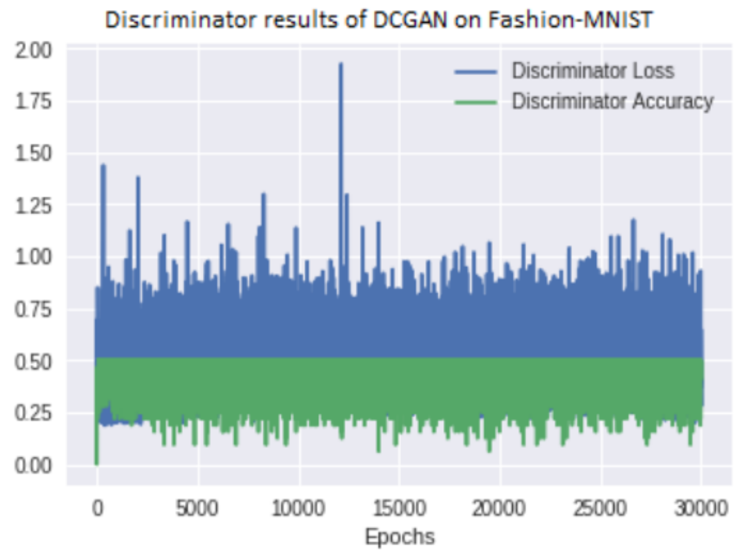

Figure 13: Loss of DCGAN model

6.2.2 CapsGAN

The below figure 14 is a graph of Loss of CapsGAN model

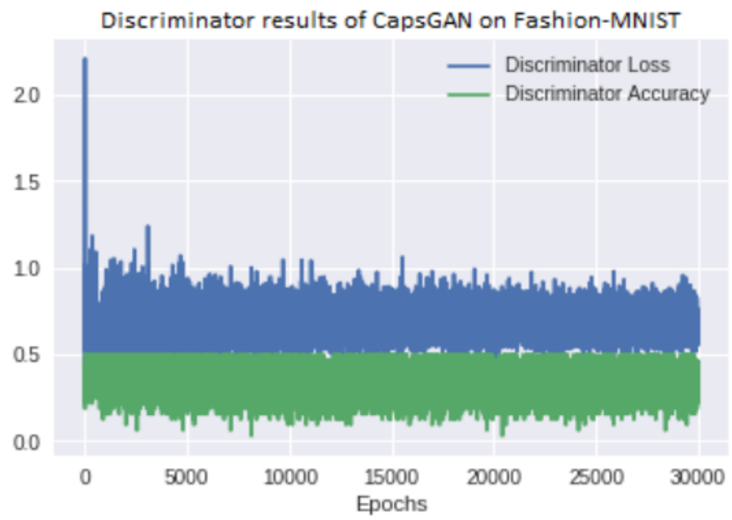

Figure 14: Loss of CapsGAN model

As seen in the graph, the loss for DCGAN was steady in beginning but reached to peak in between of 10000th – 15000th iterations. While in the CapsGAN, graph shows high loss just at beginning and slowly getting decrease while model going towards the end. Lesser the loss of model, better the performance of the model. Hence, this evaluation can be concluded as CapsGAN again have outperformed the DCGAN model as having lesser loss.

6.3 New Generated Images

The GAN models are so successful and excellent in work because of their generative nature like generating new images from the given data samples. The objective of this research was to generated new fashion outfit images for the support and alternatives to the fashion designers of Fashion Industry. This new generated images will help the fashion business if this research gets into real implementation. Below figure 15 is the new generated images by model shown as:

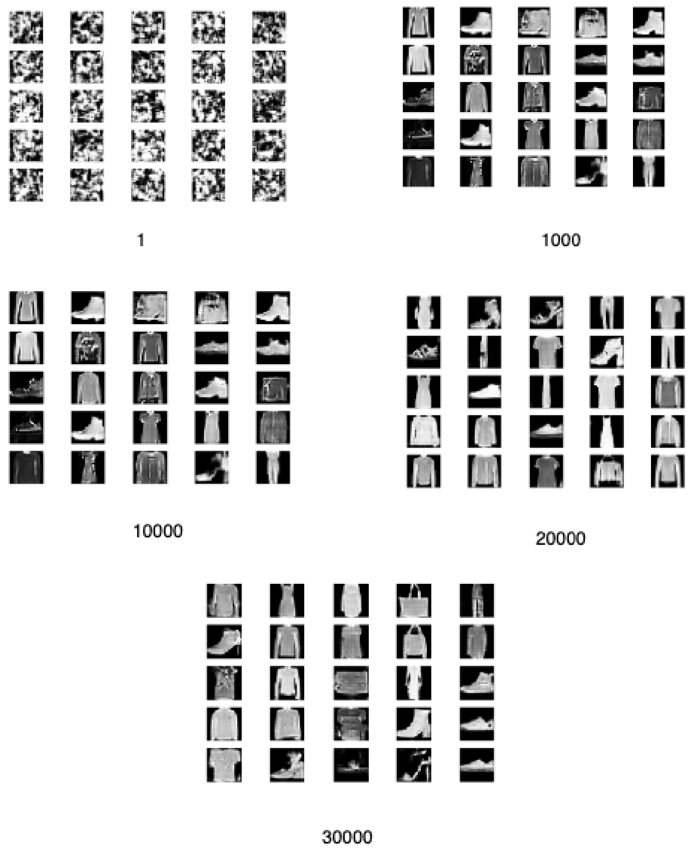

Figure 15: New Fashion Outfits Design Images generated by GAN

6.4 Discussion and Challenges

This study is done with 2 evaluation parameters of GANs. During the qualitative, it was found that CapsGAN performed better than DCGAN in the quality and precision factor. While in the quantitative side, again CapsGAN has less loss compared to DCGAN makes CapsGAN winner here. Though DCGAN was faster but CapsGAN performed better. Hence, it can be state that in comparison of both the model CapsGAN have worked more better on the classic Fashion − MNIST dataset in fashion domain. There are many challenges faced in this research like implementing basic GAN model is also difficult because of maintaining the balance between generator and discriminator is tedious. While after that implementing 2 advanced versions of GAN is more difficult. And choosing suitable dataset for that is another challenge. Moreover, neural networks takes much more time for the execution and requires high computation power to run. So, the model execution was done on Google Collaboratory having high speed GPU for lesser time as it was not possible on regular system. After all the challenges encountered, desired objective of the research was achieved.

7 Conclusion and Future Work

The main challenge that arises for the researches is not gathering enough datasets in specific fields. In order to overcome such problem GANN (Generative Adversarial Neural Networks) model have been introduced which produces new samples which in turn can be used for future research works. To enhance the effectiveness of Fashion-MNIST through this research designs had been bought in for new and specific fashion items.

Fashion-MNIST is vast and perfectly contains the need of this research work. The main purpose to carry out this research is to help the fashion designers by assisting them virtually by providing new fashion images and in doing so the whole process was fulfilled in this research. It can be guaranteed that the whole fashion industry will be benefitted if the project is having real-time implementation. Through this project the researcher also wanted to identify the most effective model in this field between CapsGAN and DCGAN. Both Qualitative and Quantitative techniques had been used in the research to evaluate the images that are generated. While Qualitative method was carried out it was found that though DCGAN was fast in the beginning but there was high superiority at the end of 30000 iterations in CapsGAN in terms of accuracy and detailing in the images. On the other hand, in Quantitative method both the Generator and Discriminator loss function values were compared so that their functioning can be evaluated. This resulted in contradiction. DCGAN had better performance for Generator loss and CapsGAN showed better performance for Discriminator. It can be found that the performance of CapsGAN was much high rated in both the cases and therefore it is considered to be more suitable model which can be used to benefit the fashion industry. The researcher also concludes that Capsule Network is better than Convolutional Neural Network. This research was not carried out for any kind of comparison but there proved to be a clear comparison between Caps Net and CNN throughout the study.

New and advanced forms of GAN are coming to the market every day. Though these are advanced further examinations can be carried out in the fashion industry in order to obtain more better and improvised results in future. Greyscale images had been used in this research as a dataset. In future complex datasets like models wearing fashion outfits in various fields like shops, fashion shows, events and others can be used. Research can also be done on smaller datasets using GAN. This research had been carried out to open the path of research in fashion world and new researches can be carried out using GAN and its advancement.

8 Acknowledgements

I would like to express special thanks of gratitude to my supervisor Dr Muhammad Iqbal for continuous support and guidance throughout the thesis. Working with such a supervisor has taught me lessons which can be used in future life. I would also like to extend my gratitude to my colleagues, classmate, friends and family who provided all facilities required and motivated me. Finally I would like to thank all the lecturers of National College of Ireland to provide knowledges throughout the course.

References

Arjovsky, M. and Bottou, L. (2017). Towards principled methods for training generative adversarial networks, *arXiv preprint arXiv:1701.04862* .

Barsim, K. S., Yang, L. and Yang, B. (2018). Selective sampling and mixture models in generative adversarial networks, *arXiv preprint arXiv:1802.01568* .

Bi, L., Feng, D., Fulham, M. and Kim, J. (2019). Improving skin lesion segmentation via stacked adversarial learning, *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1100–1103.

Chan, C. O., Lau, H. C. W. and Fan, Y. (2018). Iot data acquisition in fashion retail application: Fuzzy logic approach, *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 52–56.

Choi, J., Seo, H., Im, S. and Kang, M. (2019). Attention routing between capsules, *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 1981–1989.

Deshpande, I., Zhang, Z. and Schwing, A. (2018). Generative modeling using the sliced wasserstein distance, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3483–3491.

Eshwar, S., Rishikesh, A., Charan, N., Umadevi, V. et al. (2016). Apparel classification using convolutional neural networks, *2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, IEEE, pp. 1–5.

Fang, F., Wang, H. and Tang, P. (2018). Image captioning with word level attention, *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1278–1282.

Giri, C., Thomassey, S., Balkow, J. and Zeng, X. (2019). Forecasting new apparel sales using deep learning and nonlinear neural network regression, *2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)*, pp. 1–6.

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

Han, X., Wu, Z., Jiang, Y.-G. and Davis, L. S. (2017). Learning fashion compatibility with bidirectional lstms, *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1078–1086.

Li, J., Jia, J. and Xu, D. (2018). Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks, *2018 37th Chinese Control Conference (CCC)*, pp. 9159–9163.

Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X. (2016). Deepfashion: Powering robust clothes recognition and retrieval with rich annotations, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1096–1104.

Ma, Z., Dong, J., Zhang, Y., Long, Z., He, Y., Xue, H. and Ji, S. (2020). Fine-grained fashion similarity learning by attribute-specific embedding network, *arXiv preprint arXiv:2002.02814* .

Patel, B., Dhayal, K., Roy, S. and Shah, R. (2017). Computerized skin cancer lesion identification using the combination of clustering and entropy, *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, IEEE, pp. 46–51.

Silva, E. S., Hassani, H., Madsen, D. Ø. and Gee, L. (2019). Googling fashion: forecasting fashion consumer behaviour using google trends, *Social Sciences* **8**(4): 111.

Tang, H., Xu, D., Sebe, N. and Yan, Y. (2019). Attention-guided generative adversarial networks for unsupervised image-to-image translation, *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp. 1–8.

Vasileva, M. I., Plummer, B. A., Dusad, K., Rajpal, S., Kumar, R. and Forsyth, D. (2018). Learning type-aware embeddings for fashion compatibility, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 390–405.

Wang, R., Li, G. and Chu, D. (2019a). Capsules encoder and capsgan for image inpainting, *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, pp. 325–328.

Wang, R., Li, G. and Chu, D. (2019b). Capsules encoder and capsgan for image inpainting, *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, IEEE, pp. 325–328.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining, *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Springer-Verlag London, UK, pp. 29–39.

Xiao, H., Rasul, K. and Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, *arXiv preprint arXiv:1708.07747* .

Xu, Q., Huang, G., Yuan, Y., Guo, C., Sun, Y., Wu, F. and Weinberger, K. (2018). An empirical study on evaluation metrics of generative adversarial networks, *arXiv preprint arXiv:1806.07755* .

Yu, C., Hu, Y., Chen, Y. and Zeng, B. (2019). Personalized fashion design, *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9045–9054.

Zhai, Z. and Zhai, J. (2018). Identity-preserving conditional generative adversarial network, *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–5.

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X. and Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, *Proceedings of the IEEE international conference on computer vision*, pp. 5907–5915.