Image-Based Virtual Try-On System using Deep Learning.

By

GAN YONG HAO

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfilment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

MAY 2021

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: IMAGE-BASED VIRTUAL TRY-ON SYSTEM USING DEEP LEARNING

Academic Session: MAY 2021

I GAN YONG HAO

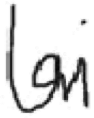
(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in
Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)



(Supervisor's signature)

Address:
NO.16 JALAN SS19/6R
47500, SUBANG JAYA, SELANGOR

Lai Siew Cheng
Supervisor's name

Date: 01-09-2021

Date: 3/9/2021

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 2/9/2021

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that GAN YONG HAO (ID No: 18ACB00330) has completed this final year project entitled "IMAGE-BASED VIRTUAL TRY-ON SYSTEM USING DEEP LEARNING" under the supervision of Ts Lai Siew Cheng (Supervisor) from the Department of COMPUTER SCIENCE, Faculty of INFORMATION AND COMMUNICATION TECHNOLOGY

I understand that University will upload softcopy of my final year project in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

GAN YONG HAO

DECLARATION OF ORIGINALITY

I declare that this report entitled "IMAGE-BASED VIRTUAL TRY-ON SYSTEM USING DEEP LEARNING" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : GAN YONG HAO

Date : 02-09-2021

ACKNOWLEDGEMENTS

First and foremost, I would like to thanks to my supervisor, Ts Lai Siew Cheng for continuous support of my final year project, for his motivation and immense knowledge. Besides, I am grateful to my batchmates, Tan De Zhian and Yong Seng Zhi who had assisted and encouraged me whenever I encounter problem in my project.

Last but not the least, I would like to thank my family and friends for supporting me spiritually throughout writing this thesis. They always comfort me whenever I am facing with difficulties and challenges. Thank you.

ABSTRACT

Due to the pandemic, there are more and more people shopping online especially for buying cloths. However, online shopping does not allow physical try-on, therefore limiting customer understanding of how a cloth will look on them. Thus, the image based virtual try-on system is introduced to allow customer to able to try on the desired cloth from online store. The aim of this project is to research and make an improvement to the existing system like CP-VTON. The system able to generate high fidelity try-on images that preserves the overall appearance and the characteristics of clothing items. Not only that, the system also able to preserve other human components other than the target clothing area by adding face, hair, lower cloths, and legs. The system consists of two modules, the Geometric Matching Module (GMM) and Try-On Module (TOM). To warp in-shop clothing item to the desired image of a person with high accuracy in GMM, grid interval consistency loss and an occlusion handling technique are proposed. Grid interval consistency loss regularizes transformation to prevent distortion of patterns in clothes and an occlusion handling technique encourages proper warping despite target bodies are covered by hair or arms. After that for TOM, face, hair, lower cloths, and legs are added to person representation input of TOM to preserve other human components other than the target clothing area. TOM will synthesize the final try-on image of the target person seamlessly with the warped clothes from GMM and the person representation. Lastly a synthesizing discriminator is also added to the end of TOM using SN-PatchGAN to further improve the quality of images generated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

%	Percent

LIST OF ABBREVIATIONS

CP-VTON	Characteristic-Preserving Image-based Virtual Try-On Network
GMM	Geometric matching module
TOM	try-on module
GIC	grid interval consistency
SN-PatchGAN	spectral normalized patch Generative Adversarial Networks
LA-VTON	Looking-Attractive Virtual Try-On
SSIM	structural similarity index
IS	Inception Score
FID	Fréchet Inception Distance
VITON	Virtual Try-On Network
GANs	Generative Adversarial Networks
TPS	thin-plate spline
RGB	Red Green Blue
WUTON	Warping U-net for a Virtual Try-On system
MG-VTON	Multi-pose Guided Virtual Try-on Network
CGAN	conditional generative adversarial network
RANSAC	Random sample consensus
ReLU	rectified linear unit
IDE	Integrated Development Environment

CHAPTER 1: INTRODUCTION

1.1 PROBLEM STATEMENT AND MOTIVATION

Due to the pandemic, more and more brick-and-mortar retail shops had moved their business online. One of the web hosting market leaders, GoDaddy had seen a very big rise in the use of their company's e-commerce product. According to GoDaddy, from February 2020 to April 202 there was a 48% increase in new paying subscribers (Schallom 2020). Not solely that, the share of online attire sales as a proportion of total fashion attire and consumer goods sales is increasing at a quicker pace compare to alternative e-commerce sector. This is because online cloth or apparel shopping able to provide the ease of shopping from the comfort of their own home, access to the newest and latest products and also a large selection of items to choose from. Unfortunately, online shopping does not allow to try-on cloth physically, therefore limiting the consumer comprehension of how the cloth will look on them. This essential limitation inspired the development of virtual try on rooms, where pictures of a client wearing their desired cloth are produce synthetically to help choose and compare the most desired look.

Recently there are more and more useful clothes datasets for training like DeepFashion (Ziwei Liu, 2016). Because of that, the huge number of cloths datasets from internet give the opportunity to have the virtual try-on task with the use of 2D modeling. Early virtual try-on systems rely on 3D information such as the garment and even the body shape, this kind of 3D information is not possible for normal users and everyday used. Not only that 3D information is expensive and take time as the body shape and garment needed to be scanned to produce as 3D model.

Furthermore, recent image synthesis method for virtual try-on fail to retain small details, such as the person's hair and the lower-body clothing, which lose their details and style. To create a image that is realistic, such technique use a coarse-to-fine network to create an image that is only conditioned on clothing. They overlook important aspects of human parsing, resulting in a fuzzy and irrational picture synthesised. Thus, it prompts us to make this system to tackle the problem. Also most of them unable to handle occlusion like hand or hair blocking the body. Other than that, some of the system have issue with preserving the cloth detail like logo, text, and hair.

1.2 PROJECT SCOPE

The final product of this project, a improve plan for the existing system is introduced. This improves system able to overlays new clothes onto the person body and it preserves and enhances rich details in salient regions like facial identity, the hair of the person and clothing of lower body lose the details and style without resorting the use of 3D information. It will be a website where user able to transfer the in-shop clothing image to the target user photo without the loss of facial detail and distortion of patterns and prints. This system is a improve over the existing system which is CP-VTON. Which use Geometric matching module (GMM) to transform the cloth user want to wear into warped cloths which is roughly aligned with user body. The second part is try-on module (TOM) which fuse the warped clothes with the user image (Thibaut et al., 2019).

This system is programmed using python programming language on PyCharm community edition with the use of Pytorch version 1.8.0 which provide library for deep learning that needed to perform task like geometric matching. Inside the system, there are few techniques used. One of it is person representation which contains a set of features like human body representation, face and hair segment and pose heatmap. the second is GMM, which do occlusion handling and grid interval consistency (GIC) loss to prevent high deformation of cloth prints and patterns. The third technique is TOM, Inside TOM it will have mask and person image generator that is based on U-net and it also has a synthesizing discriminator that is based on SN-PatchGAN during training.

1.3 PROJECT OBJECTIVES

This project aims to help online shopper to try out the cloths by transfer the clothes from in-shop clothing image to the target user photo and it is based on CP-VTON. This project objectives are as follows:

- Develop a web application where user able to input in-shop clothing image they wanted to wear, user photo and select pose they wanted. Then the system will output an image with the user wearing the in-shop clothing and selected pose.
- Implement an occlusion handling that able to generate precise geometric transformation nor matter of occlusion by arms and hair. The system able to wrap the shirt perfectly to the person body, while the body is being block by hand or hair.
- Implement a grid interval consistency that retain the feature of clothing product, such as logo, text and texture without any perceptible distortions and artifacts. The system able to product the person image wearing the desired shirt without the loss of detail from the shirt like logo, text and texture
- The system able to preserve other human features other than the target clothing area by add on face, lower cloths, hair, and legs to human representation as an input in TOM. Other human features from the final image produce should not be loss or different from the original image.

1.4 IMPACT, SIGNIFICANCE AND CONTRIBUTION

This virtual try on system allow users to experience themselves wearing different clothes without efforts of changing them physically. This comes in handy as online shopping become popular because of the pandemic. Where consumers are concerned about how a particular cloth in a product image would look on them when buying apparel online. So, this virtual try on system allow customer to virtual try on cloth will definitely enhance their shopping experience and transforming the way people shop for cloths, it also allows retailers to save some cost. Not only that, but this system also helps customers to quickly judge whether they like the cloth or not and make buying decisions and improves sales efficiency. For example, when a customer wanted to buy a cloth online through a retailer website, he/she will prefer shopping at website with virtual try on system as it allows them to quickly decide whether the cloths look good on them or not. Compare to a retailer website without virtual try on system, customer need to guess whether they like it or not. This will cause them to hesitate to buy the cloth.

But the existing virtual try on system unable to produce an image that is desirable like most of the have problem like unable to preserve the characteristic of the cloth, some is they cannot system ignore the preservation of other body component like pants or legs. Most of all system are unable to handle occlusion like the cloth being block by hair and arms. Thus, to handle the problem a few techniques is use like Grid Interval Consistency and Occlusion handling to preserve the characteristic of the cloth, so it will not over deform.

1.5 BACKGROUND INFORMATION

Virtual try on system can provid**e** information about the product, similar to the information obtained by directly testing the product. In addition, the interactivity and customer l**o**ya**l**ty **g**ene**r**ated by the virtual try on system can increase the entertainment value of online shopping. The traditional way for synthesizing realistic picture of individuals sporting cloths rely on detailed 3D models engineered from either depth cameras or multiple 2D images. 3D models allow realistic clothing simulation under geometric and physical constraints, as well as control of the viewing direction, lighting, texture and pose. However, they cost a lot in terms of data capture, annotation, computation and in some cases the need for specialized devices, such as 3D sensors. These giant prices hinder scaling to numerous customers and garments. One of the

example systems that uses this technique is virtual fitting by single-shot body shape estimation by Masahiro (Masahiro et al., 2014).

Therefore, pure image-based methods have been proposed, this deforms clothes from clean product images by applying the geometric transformation and then fit the deformed cloths to the original person image. With this it provides a more economical solution as it does not require sensor to get the 3D information. With the current advance and research in the conditional image generation like image-to-image translation make it suitable for the virtual try on system. But this kind of technique produce blurry image and misses critical details. Till now the best practice for image-conditional virtual try-on is still a two-stage pipeline that introduce by VITON (Han et al., 2018). But VITON performances are far from the plausible and desired generation. So, to address the problem, CP-VTON is introduce. Where Characteristic-Preserving Image-based Virtual Try-On Network (CP-VTON) able to retain the shirt characteristics, such as logo, texture, and text. They introduce a new learnable thin-plate spline transformation using a tailored convolutional neural network so that the cloths user wanted able to align with target person (Wang et al., 2018).

.

1.6 PROPOSED APPROACH

This is a web application where user able to input desired cloth image and person image, then the system will output the person image wearing the desired cloth. This system is inspired by CP-VTON+ (Matiur et al., 2020) and LA-VTON (Lee et al., 2019). The system adopts use a strategy call outline-course-fine which divide the task into a three subtask which is person representation, Geometric Matching Module (GMM) and Try-on module (TOM). The main purpose of GMM is align the desired shirt to the person body and output a warped cloth. Inside GMM there is grid interval consistency and occlusion handling to warp the desired cloth more precisely to the person. After that, the warp cloth will go to TOM which take in the warp cloth and produce the final try-on image with the person wearing the desired cloth. In TOM there is synthesizing discriminator using SN-PatchGAN to further improve the quality of pictures generated. After this, the image will be display on the web page for the user to see.

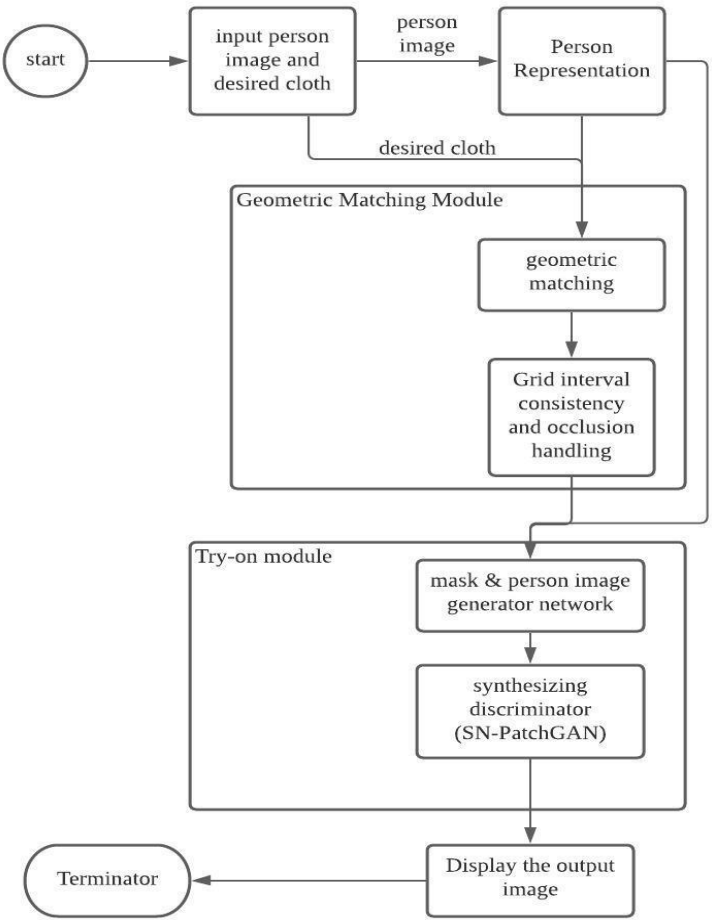


Figure 1.6.1: System overview

1.7 REPORT ORGANIZATION

There are five different chapters in this report; Chapter1 Introduction, Chapter 2 Literature Review, Chapter 3 System Design, Chapter 4 System Implementation, Testing, and Result, and Chapter 5 Conclusion. Beginning with the introduction, which includes the problem statement, motivation, project scope, project objective, background information, proposed approach, and report organization. The second chapter begins with a review of the literature. The design of the system is laid out in Chapter 3 to describe how to develop this project. This chapter will include a diagram of the system's top-down design. The implementation of the system and the experiment results will be covered in Chapter 4. The final chapter will provide a conclusion to the entire project.

CHAPTER 2: LITERATURE REVIEW

2.1.1 Virtual Fitting by Single-shot Body Shape Estimation (Masahiro et al., 2014)

Masahiro et al. proposes an original virtual fitting system, with the objective of adjusting 2-dimensional clothing images to models. This is executed using their 3-dimensional models from single-shot depth images. Then, the system takes advantage of a method that overlays clotting images obtained from a person given their body shape.

Firstly, the input data being processed are the image of the body and the depth image, both acquired at the same time. Next, the input data is utilized when estimating single-shot body shapes. These body shapes estimations compare pre-trained depth images with the input depth image, finally determining a three-dimensional body shape model accurate to the user. Then, the processing step after that is the selection of clothing images. This processing step takes the 3D body shape from a single shot body shape estimate as an input and scavenges for body shape models that are similar in the clothing database. The end result of this process is a selected image taken from the database of clothing with the manifold shape of the body. Finally, the last processing step is the transformation of clothing images. Looking at things from this perspective, this step takes the body image, depth image, the three-dimensional body shape model estimates and also the selected image of the clothing to output a final overlaid image.

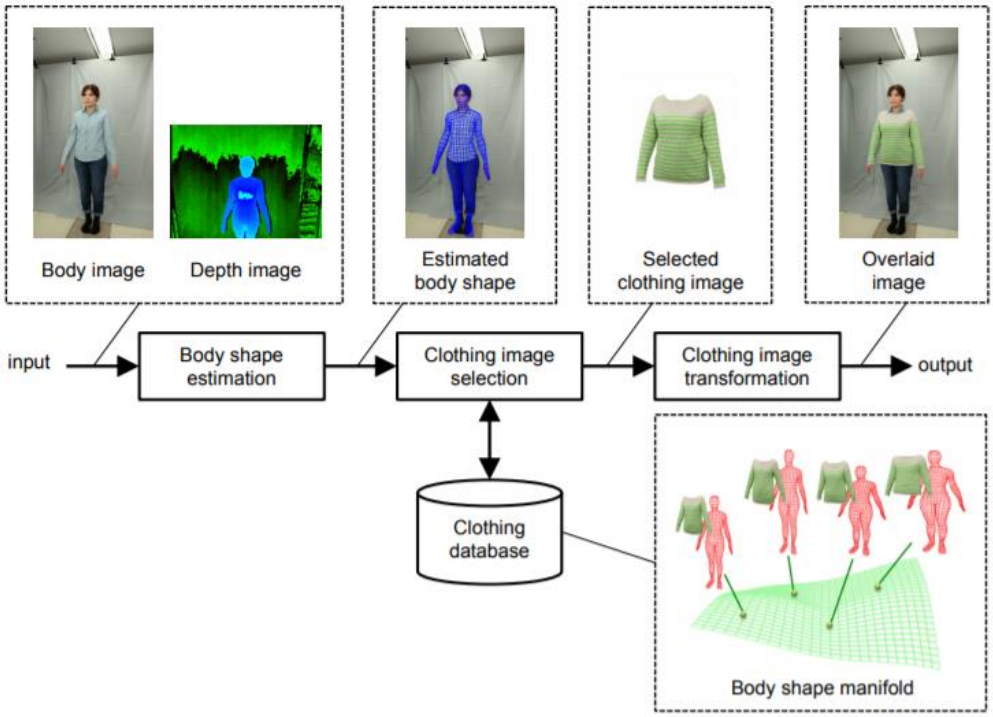


Figure 2.1.1.1: proposed process flow of the system

Overall, this system is able to produce very realistic simulations of clothing even under constraints in aspects such as geometrically and physically. Also, it is worth nothing that this system has near perfect control of the direction of view and also texture. Contrary to that, the disadvantage is that it bears huge costs. This is because the process involves capturing of data, annotation, computational resources and also specialized devices like depth sensors. These huge costs often obstruct the scalability of the system to the masses of customers and garments.

2.1.2 VITON (Han et al., 2018)

Han et al. propose VITON that able to seamlessly shift the target clothing article of a product image to an accurate region of a clothed person in a two-dimensional image with the use of Conditional Generative Adversarial Networks (GANs). First, they introduced a clothing-agnostic person representation which comprised of several features that were utilized to describe certain characteristics of a person. This feature extracts the facial, body shape, hair and even the pose given a reference image.

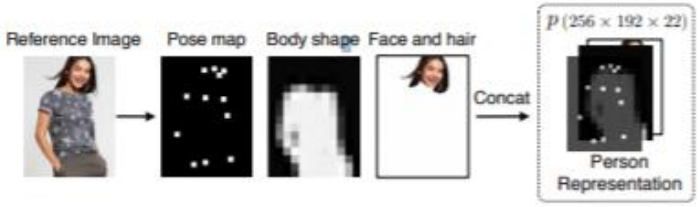


Figure 2.1.2.1: A clothing-agnostic person

After that they use the information from the clothing-agnostic representation as part of input for the generator in the next section which is multi-task encoder-decoder generator. They used a multi-task encoder-decoder generator for the purpose of generating a coarse and synthetically clothed person. This person is made to be in the same pose, wearing the same target clothing, and also has a corresponding mask of the clothing region. Next, the clothing mask and target clothing image is used in shape and context matching for the estimation of thin-plate spline (TPS) transformation and ultimately, to generate a warped image of the clothing.

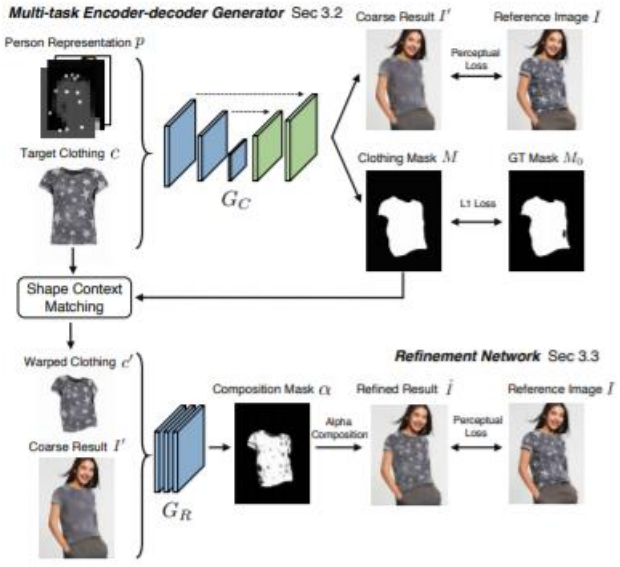


Figure 2.1.2.2: An overview of VITON

Moreover, they take advantage of a refinement network that is thoroughly trained for the objective of learning how to properly composite a warped clothing to a coarse image. This is

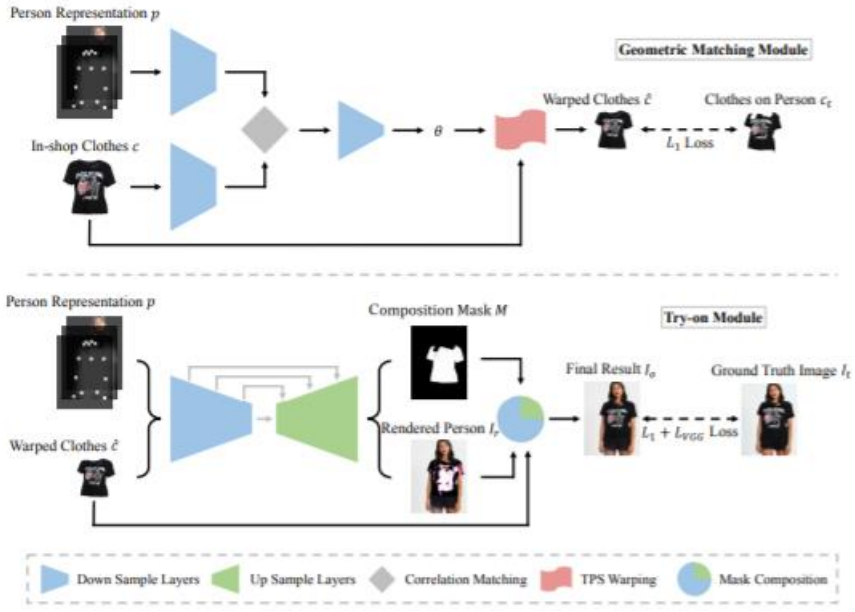
done so that the desired item is transferred naturally without deformations, and with a rather detailed visual patter. The key feature of VITON can shift an item of clothing in a product image to a person, mainly relying on RGB images. The limitation is that VITON is imperfect in terms of retaining the characteristics of cloth such as their texture and logo. (Han et al., 2018).

2.1.3 CP-VTON (Wang et al., 2018)

Wang et al. introduced Characteristic-Preserving Image-based Virtual Try-On Network (CP-VTON) that is able to achieve a convincing try-on image syntheses. Compared to other networks, it is also able to preserve the characteristics of cloth better as well. CP-VTON is separate to two parts, namely the Geometric matching module (GMM) and try-on module (TOM).

First part is Geometric Matching Module (GMM) which is used to transform the clothes of the target into warped clothed. Then, it is aligned with a representation of the input person. Furthermore, in GMM, there is a proposal of a new learnable thin-plate spline (TPS) transformation. To elaborate, it is done through a tailored convolutional neural network to accurately line up the clothes inside the store with the target's representation. Also, the network parameters are trained from paired images consisting of clothes in the shop and also a wearer. This is done without obtaining any requiring any explicit correspondences of interest points.

The next part is try-on module (TOM) which fuses the warped clothes and the target, and then synthesizing the final result. In TOM, the module will take the warped clothes produce in GMM and person representation as an input and generates the rendered person's image. Following that is a composition mask that delineated the aligned clothes' details, which is kept in the synthesized image. Lastly, the rendered person and the warped clothes are further fused together using the composition mask to create the final try-on result.



Fiaure 2.1.3.1: Overview of CP-VTON

In conclusion CP-VTON able to achieve a very convincing try-on image syntheses, and simultaneously saving the characteristics of cloth. However, the limitation of this system lies within its improperly preserved shapes of old clothing and alson unusual poses. (Wang et al., 2018).

2.1.4 WUTON (Thibaut et al., 2019)

Thibaut et al. proposed a system named WUTON. It stands for a Warping U-net for a Virtual Try-On system. The strength of this system is that it is able to eliminate the need of a final composition step which is largely found in the current best approaches, mostly done with an adversarial trained generator. Other than that, this method performs much better on the edges of the replaced objects, and also offers a more natural appearance of shadows and contrasts. The system they proposed consists of two connected modules. Firstly, the convolutional geometric matcher, similar to CP-VTON. This module is made up of two feature extractors, named F1 and F2. These extractors are standard convolutional neural networks as shown in Figure: 2.1.3.1. It can be seen that the correlation map C is able to catch dependencies even of distant places of two distinct feature maps. This is useful for the purpose of aligning two images. Also, the input of the regression network, C, is used to generate the output parameters θ, allowing it to perform geometric transformations Tθ. Furthermore, they also use TPS transformations, allowing them to generate relatively smooth sampling grids when provided control points. Given that they are transforming deep feature maps of the U-net generator, they are able to create a sampling grid for every U-net scale given parameters θ.

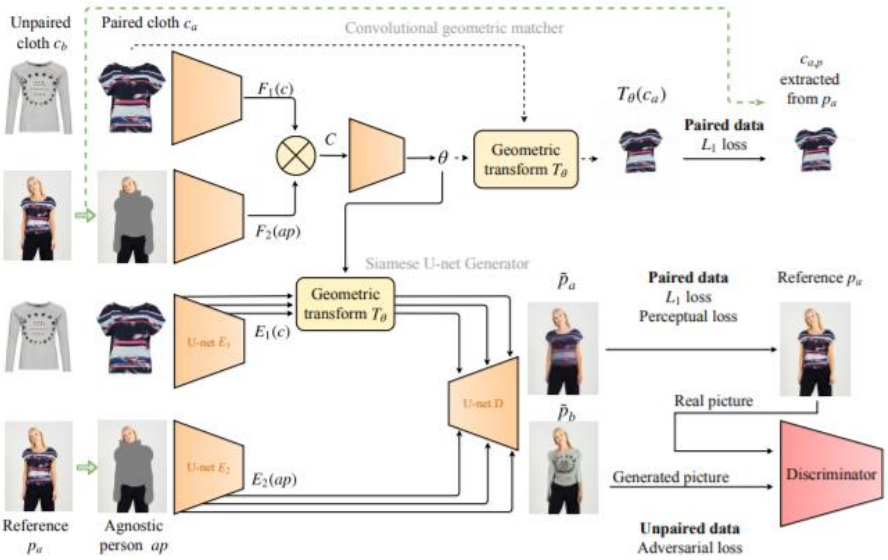


Figure 2.1.4.1: Overview of WUTON

It is also observed that the second model is indeed a U-net generator with the objective to synthesize the output image. The input consists of a few tuples of images. These two pictures are not aligned in the spatial dimension and cannot be easily concatenated to feed a standard U-net. To solve this problem, two different encoders named E1 and E2 were used. These

encoders process the images independent of each other, and even using independent parameters. From this point, the feature maps of E1 and E2 are then concatenated together to feed the decoder. Utilising these feature maps which are already aligned, the generator is capable of composing them to produce more real results. Given the reason that they attempt to concatenate the feature maps and using the U-net decoder for composition, experiments will definitely show that it exerts more flexibility and also capable of producing natural end results.

In conclusion, this system able to remove the requirements of the step involving the final composition in the current best approaches using the adversarial trained generator. This allows them to generate a more natural and detailed synthesis. But the weakness of this system is that some of the body detail like hand have been distorted.

2.1.5 MG-VTON (Dong et al., 2019)

Dong et al. introduced a system which is MG-VTON that able to map an input image of a model to a different image of the same model, but this time with a new outfit and a unique pose. This is done by manipulating the clothing and pose of the target. MG-VTON consist of three stage. Stage 1 which is conditional parsing learning, first they decompose the image of reference into three separate binary masks, face mask, and body shape. Then, they are joined together with the target cloths' image. The target pose is also used as an input for the conditional parsing network, used to estimate the human parsing map. The conditional parsing network is based on the conditional generative adversarial network (CGAN, aimed to create convincing results from manipulation of images. After that for Stage 2, the clothes were warped and removed from the image of reference. They are then further concatenated with the target pose, and then synthesized parsing for the coarse result. This uses the technology called Warping Generative Adversarial Network (Warp-GAN). This network is aimed at synthesizing the coarse result, curbing the alignment problem caused by the abundant poses and clothing. In the third stage, the coarse result is refined using a refinement render. The purpose of this stage is to recover important details such as texture and to ease the artifacts that were there because of the alignment issues of the reference and target pose.

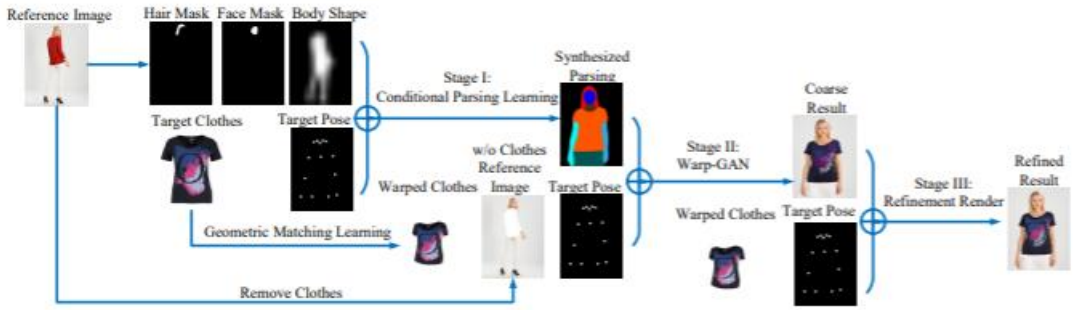


Figure 2.1.5.1: Overview of MG-VTON

Other than that, this paper also introduced a new dataset for this experiment which name MPV. MPV consists of over 35,000 images of people and over 13,000 images of clothing. Each and every of the people images in MPV has distinct poses. It is worth noting that the resolution of the image is rather small, which is 256 by 192. The authors further took out approximately 52,000 three-tuples of the same person wearing the same clothing, but the difference is they have various poses. To elaborate further, these images are then separated into training sets and test sets, consisting of about 52,000 and 10,500 three-tuples respectively. When compared to

datasets such as DeepFashion which only has pairs of the same person in different poses, they are superior in the fact that they have image of clothes (Ziwei Liu, 2016).

In conclusion, MG-VTON able to generate a new person image after properly aligning the clothes of choice onto the image of the input person and manipulating human poses. But there is some limitation of this system which is that most of the face have been distorted.

2.1.6 LA-VITON (Lee et al., 2019)

Lee et al. introduce architecture is like CP-VTON as they also use the GMM and TOM as well. They added, Grid Interval Consistency and Occlusion Handing. They use grid interval consistency because TPS transformation in the GMM is highly flexible often cause the patterns and print to be distorted. This is also observed when coarse-to-fine strategy is applied. When providing TPS a high degree of freedom, it often times result in unwanted warping. Thus, the grid interval consistency loss was introduced, retaining details of cloths even after the warping process. This retaining was not only limited to patterns of the clothes, but also the shapes, keeping consistent during the intervals.

Figure 2.1.6.1: comparison of the warp cloth with GIC loss and without (Lee et al., 2019).

While for the occlusion handing, they needed occlusion handing because person can be very easily blocked by body parts like limbs or hair. Furthermore, since they extract the cloths region using ground truth named Look-Into-Person, this network then attempts to fit the clothes in the shop onto the cloths which could be observed and extracted. This problem is then solved by not including the obstructed regions from warp loss calculation. It is lenient in network training, allowing it to be thoroughly trained for better results when transforming parameter estimation. So, with this, this network is capable of performing geometric transformation without needing to care about hindrances such as body limbs or hair.(Lee et al., 2019).

Figure 2.1.6.2: comparison of warp with occlusion handling and without (Lee et al., 2019).

2.1.7 CP-VTON+ (Matiur et al., 2020)

CP-VTON+ is built on the architecture structure of CP-VTON as well. There are several places they improve the architecture. First is in the GMM stage, they relabel all the VITON dataset as bare chest area and the neck is wrongly labeled as background and sometime the shape of the body will be deform wrongly by hair occlusion. Secondly, they also replace the colored texture from try-on clothing as it does not help in matching process instead, they use clothing mask. The last change for the GMM network is that they added regularization for the TPS parameters. They do the regularization is because the existing methods reveal that warped clothing is often severely distorted. The new TPS parameters defined on the grid deformation and not directly on the TPS parameters.

$$L_{GMM}^{CP-VTON+} = \lambda_1 \cdot L1(C_{warped}, I_{C_t}) + \lambda_{reg} \cdot L_{reg} \quad (2)$$

$$L_{reg}(G_x, G_y) = \sum_{i=-1,1} \sum_x \sum_y |G_x(x+i, y) - G_x(x, y)| + \sum_{j=-1,1} \sum_x \sum_y |G_x(x, y+j) - G_x(x, y)| \quad (3)$$

Figure 1.1.7.1: overview of grid warping regularization

For the TOM stage, they added the face, lower cloths, hair, and legs to the human representation input of TOM. They do this to keep the other human feature other than the target clothing area. They also added the binary mask of warped clothing to the TOM network input, they do this because the TOM unable to recognize the white clothing area as being the same as the in-shop clothing image background (Matiur et al., 2020).

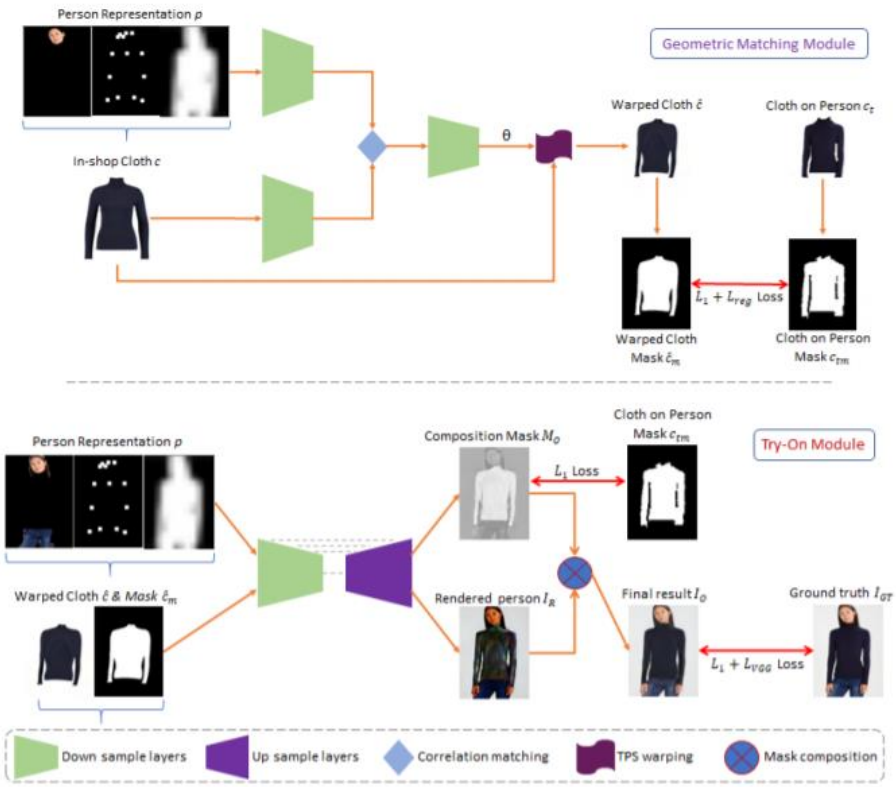


Figure 2.1.7.2: overview of CP-VTON+ pipeline

2.2 COMPARISON OF FEATURES

Each system had their own unique features and also had essential functions for the users as well. The comparison of features between the system are as follows:

Table 2.2.1: Comparison of features

Features / Methods	VITON	CP-VTON	WUTON	MG-VTON	Single-shot Body Shape Estimation	CP-VTON+	LA-VTON
3D modelling technique					yes		
image-based virtual try-on	Yes	yes	yes	yes		yes	yes
preserving cloth characteristics		Yes	yes	yes	yes	yes	yes
manipulating human poses				yes			
Realistic deformation to preserve complex patterns			yes				yes

Occlusion handling				yes			yes
Preserve other human feature beside the target clothing area						yes	

Based on the Table 2.2.1 all the system except Single-shot Body Shape Estimation able to do image-based virtual try-on. As image-based virtual try-on is batter because the cost of developing for 3D modelling technique is higher as data capture, annotation, computation and in some cases the need for specialized devices, such as 3D sensors are very expensive. Moreover, only VITON unable to preserving cloth features, such as text, logo and texture. Furthermore, WUTON able to do realistic deformation by using U-net to preserve complex patterns like stripes. Besides that, only MG-VTON able to manipulate human poses of the final image. Other than that, CP-VTON+ able to preserve other human features other than the target clothing like face, hair, lower cloths, and legs. In additions, the TPS transformation use in VITON, CP-VTON and MG-VTON generally shows good performance, but its high flexibility often causes distortion of patterns and prints. Thus LA-VTON is introduce as it able to perform realistic deformation using grid interval consistency and occlusion handing. But the only downside of the LA-VTON is it unable to preserve other human features other than the target clothing area. So, to further improve we develop a system that able to preserve other human components, perform realistic deformation and do occlusion handing is introduced by combining the LA-VTON and CP-VTON+.

2.3 Overview of Geometric Matching Module (GMM)

The GMM used in most of the system are based on the paper published by Rocco et al. in 2017 the title of the paper is "Convolutional neural network architecture for geometric matching" In the paper the parameter of a geometric transformation between two input images is estimated using a new convolutional neural network architecture. This architecture is design to mimic the classical computer vision pipeline which consists of different stages. At the first stage both input image is extracted by using a local descriptor. The second stages are the descriptors are matched across images to form a set of tentative correspondence. The last stages are estimate the parameters if the geometric model using RANSAC or Hough voting.

The architecture they introduce try to reproduce the same process which use a convolutional layer that take in the input image I_A and I_B. the use of convolutional layer is to extract feature maps f_A and f_B which are analogous to dense local descriptors. Next is matching the feature maps across images into a tentative correspondence map f_{AB}. After that, the parameters of the geometric model, θ, in a robust manner is outputted by a regression network. With this GMM, the geometry estimation tasks can be trainable end-to-end (Rocco et al.,2017).

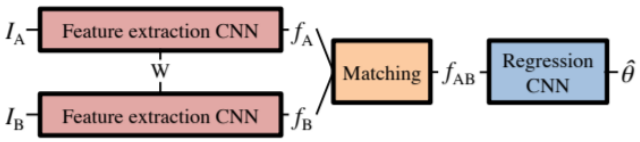


Figure 2.3.1: Diagram of the proposed GMM architecture

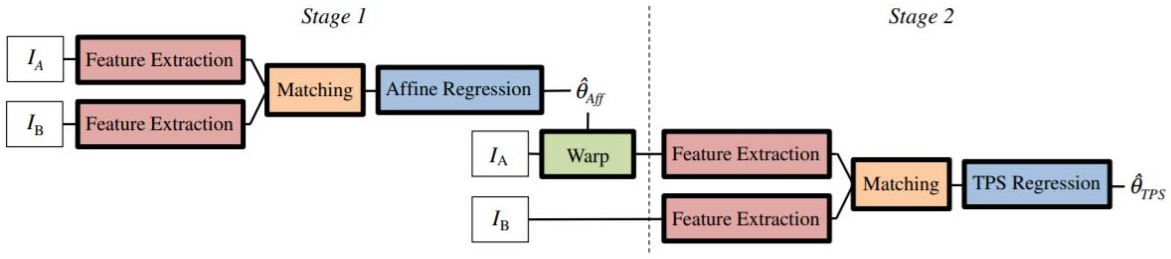


Figure 2.3.2: A stage two is introduced which contain similar thing with stage one but with thin-plate spline (TPS) regressions (Rocco et al.,2017).

Because of that, to change the desired cloths to warped cloths a new GMM is used in our system. The new GMM consist of four section. The first section is to extract the high-level features of person representation, two networks is introduced. The second section is to mix two features into one tensor, a correlation layer is introduced. The third section is to predict the spatial transformation parameters, the one tensor in correlation layer is used in the regression network. Last section is to warp an image into the output using a TPS. The pixel-wise L1 loss is counted using the different of the warped cloth and ground truth.

2.4 Overview of Try-on Module (TOM)

The main use of the try-on module is to combine the wrap cloths with the target person. The try-on module combines two technique. The first is to preserve the characteristics of warped cloths, the warped will be directly pasting onto the target person image. But the downside of this is it will produce unnatural appearance at the boundary regions of cloths and some part of the body will undesired occluded. Another technique is using U-net which introduce by Olaf et al. in 2015.

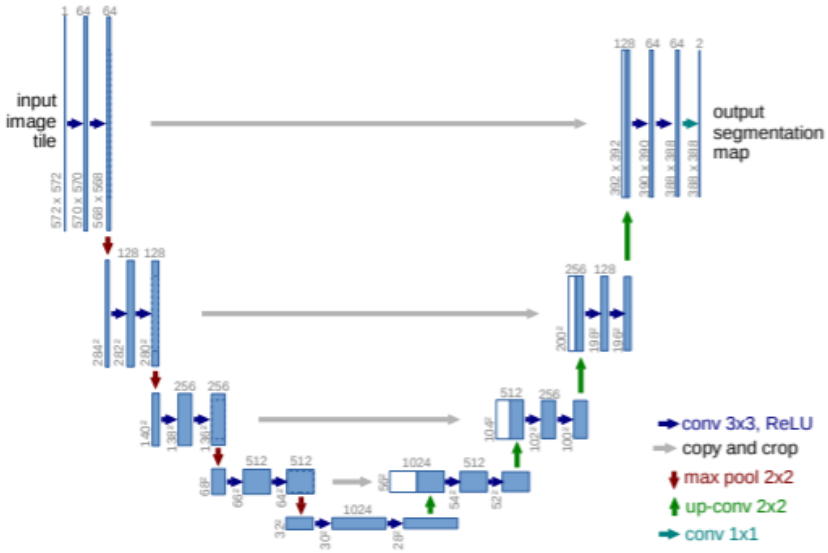


Figure 2.4.1: U-net architecture

The architecture of U-net is consisting of a contracting path (left side) and an expansive path (right side. For expansive path it consists of an up sampling of the feature map followed by a 2x2 convolution (“up-convolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. While the contracting path consists of the repeated application of two 3x3 convolutions, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for down sampling. At the bottom final layer, a 1x1 convolution is used to map each 64- component feature vector to the desired number of classes. So, it has a total of 23 convolutional layers. However, there is some drawback for applying U-net is that U-net lacks explicit spatial deformation ability, even minor misalignment could make the U-net rendered output blurry (Olaf et al., 2015). Thus, TOM use technique to make the output image look good. Because of that, a new TOM use in our system. Which get the person representation and warped cloth as an input. Then it used a U-net to predict a

composition mask and a person image is rendered. Lastly, the warp cloth and the rendered person will be combine with composition mask to generate person wearing the cloth.

CHAPTER 3: PROPOSED METHOD/ APPROACH

3.1 DESIGN SPECIFICATIONS

3.1.1 METHODOLOGIES AND GENERAL WORK PROCEDURES

This is a web application where user able to input desired cloth image and person image, then the system will output the person image wearing the desired cloth. This system is inspired by CP-VTON+ (Matiur et al., 2020) and LA-VTON (Lee et al., 2019). The system adopts use a strategy call outline-course-fine which divide the task into a three subtask which is person representation, Geometric Matching Module (GMM) and Try-on module (TOM). The main purpose of GMM is align the desired cloth to the person body and output a warped cloth. Inside GMM there is grid interval consistency and occlusion handling to warp the desired cloth more precisely to the person. After that, the warp cloth will go to TOM which take in the warp cloth and produce the final try-on image with the person wearing the desired cloth. In TOM there is synthesizing discriminator using SN-PatchGAN to further improve the quality of pictures generated. After this, the image will be display on the web page for the user to see.

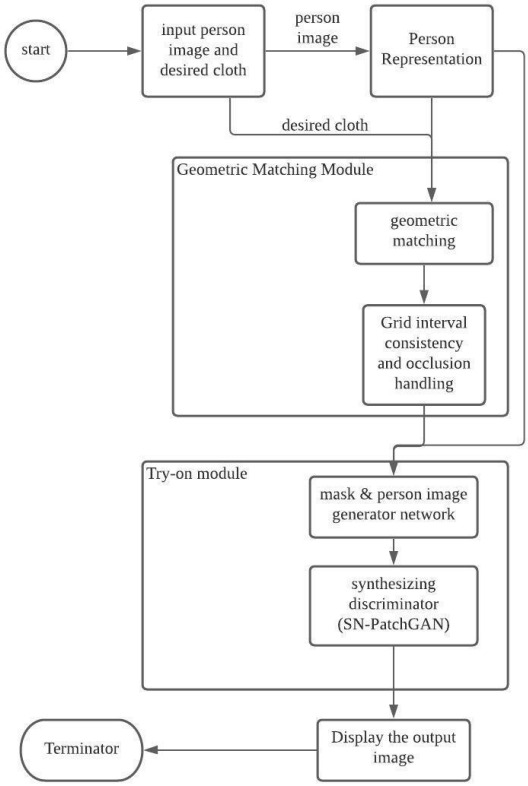


Figure 3.1.1.1: System overview

3.1.2 TOOLS TO USE

Hardware

The hardware that requires for developing this system is:

1. Laptop

Table 3-1-2-1: Laptop Hardware Specification

Operating System	Windows 10 64-bit
CPU	Intel Core i5-8250U @ 1.60GHz
Display Size	15.6” FHD (1920 x 1080)
Memory (RAM)	12GB
Storage	1TB SATA 5400RPM HDD
Graphic Card	NVIDIA GeForce MX150 2GB DDE3

Software

The software that has been chosen for developing this system are:

- Visual Studio Code

  Integrated Development Environment (IDE) is used to develop the system.
- Anaconda

  environment management system that installs, runs, and updates packages.
- Python Library Packages:
  - PyTorch Library
  - CuPy Library
  - OpenCV Library
  - Scikit-image Library
  - Scipy Library
  - Pillow Library
  - Mediapipe Library
  - Flask Library

3.1.3 SYSTEM PERFORMANCE DEFINITION

The system should have better performance than the existing system like CP-VTON+ and CP-VTON by comparing Fréchet Inception Distance (FID) score.

3.1.3.1 Qualitative Analysis

Comparison of the result form the existing system by finding the differences from the image generated and compare the image generated with other existing system.

3.1.3.2 Quantitative Analysis

To show the numerical comparison between the system the Structural Similarity Index (SSIM) are often used for the same clothing retry-on cases like when there is ground truths image. SSIM are for the warping stage and the blending stage. The SSIM is used because it is a perceptual metric that quantifies image quality degradation that caused by processing (Zhou et al., 2004). Other than that, for different clothing try-on where there is no ground truth. The Inception Score (IS) is used as it is a metric for automatically evaluating the quality of image generative models (Salimans et al. 2016). But IS and SSIM is not used as applying the IS and SSIM is only suitable to model train on ImageNet dataset which is not the datasets the system is train on. Thus, the Fréchet inception distance (FID) is used as it able to compares the distribution of generated images with the distribution of real images that were used to train the generator (Heusel et al. 2017). FID is a measure of similarity between two datasets of images. It has show that FID correlate well with human judgement of visual quality. FID is mainly used to evaluate the quality of Generative Adversarial Networks.

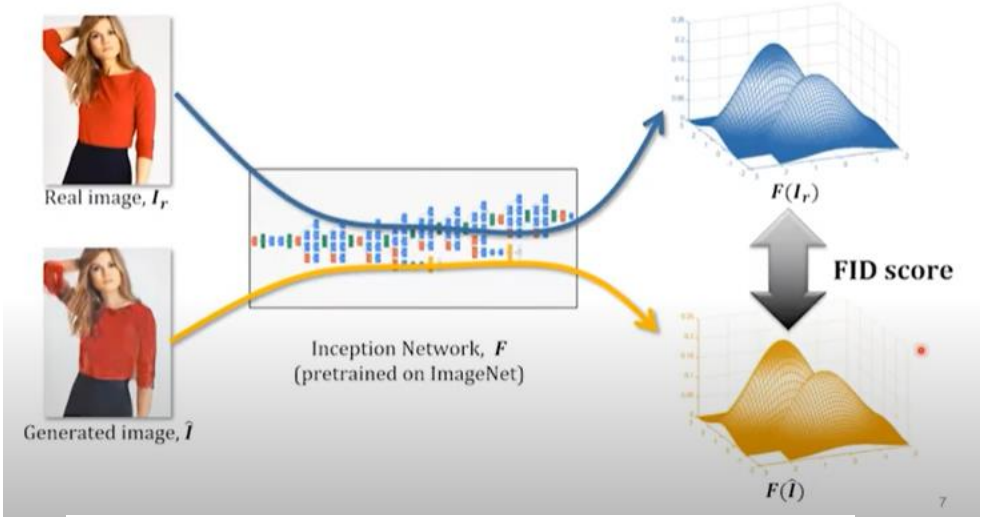


Figure 3.1.3.2.1: Overview of FID (Martin et al. 2018)

3.1.4 VERIFICATION PLAN

The system should be able to do occlusion handling where the system able to generate a very precise geometric transformation nor matter of occlusion by hair and arms. Not only that, but the system should also be able to retain the feature of clothing items, such as logo, texture and text, without any clear distortions and artifacts. Lastly it also able to preserve other human features other than the target clothing area.

A) The system should be able to handle cloth image with complex pattern and the final output image should be able to preserve complex pattern of the cloth.

Figure 3.1.4.1: Example of complex pattern cloth

B) The system should be able to do occlusion handling where it able to produce a precise geometric transformation nor matter of occlusion by arms and hair.

Figure 3.1.4.2: example of occluded image by hair

C) The system should be able to preserve other human features other than the target clothing area like face, hair, lower cloths, and legs.

Figure 3.1.4.3: example of image with pants

3.2 SYSTEM DESIGN/ OVERVIEW

The system proposed able to learns to reproduce a new person image from virtual try-on by manipulating cloths, given an input person image and a cloth image that the person wanted to wear. The system that we propose aims to produce a picture where the person is wearing the cloth they wanted. This system is inspired by CP-VTON+ (Matiur et al., 2020) and LA-VTON (Lee et al., 2019). The system uses a strategy call outline-course-fine which separate the task into a three subtask which is person representation, Geometric Matching Module (GMM) and Try-on module (TOM). The use of GMM is align the desired cloth to the person body and output a warped cloth. Inside GMM there is grid interval consistency and occlusion handling to warp the desired cloth more precisely to the person. After that, the warp cloth will go to TOM which take in the warp cloth and person representation to combine it to produce the person wearing the desired cloth. In TOM there is synthesizing discriminator using SN-PatchGAN to further improve the quality of images generated.

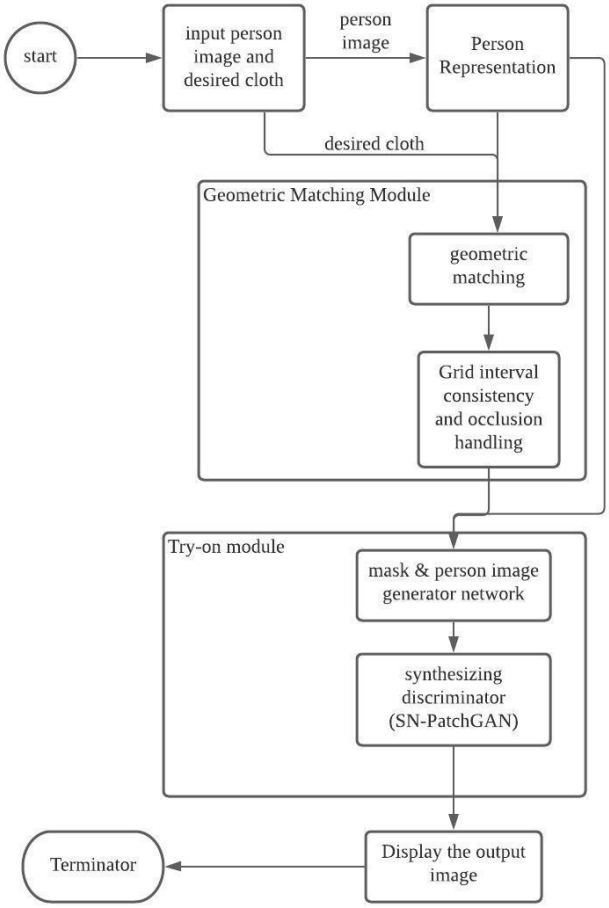


Figure 3.2.0.1: System overview

3.2.1 Dataset

The dataset use is collected from VTON (Han et al., 2018). Which contain around 19000 front-view of women and top clothing image pairs. Form the 19000 cloths, 16253 of them are usable pairs, which are then split into training set and a testing set. The training set contain 14221 pairs of images and for the testing set contain 2032 pairs if image. The dataset also contains cloth mask, human pose key point and person parse. Also, the testing set is rearranged into unpaired pairs for testing. figure 3.2.1.1 , figure 3.2.1.2, figure 3.2.1.3 and figure 3.2.1.4 is the training image from the VTON datasets.

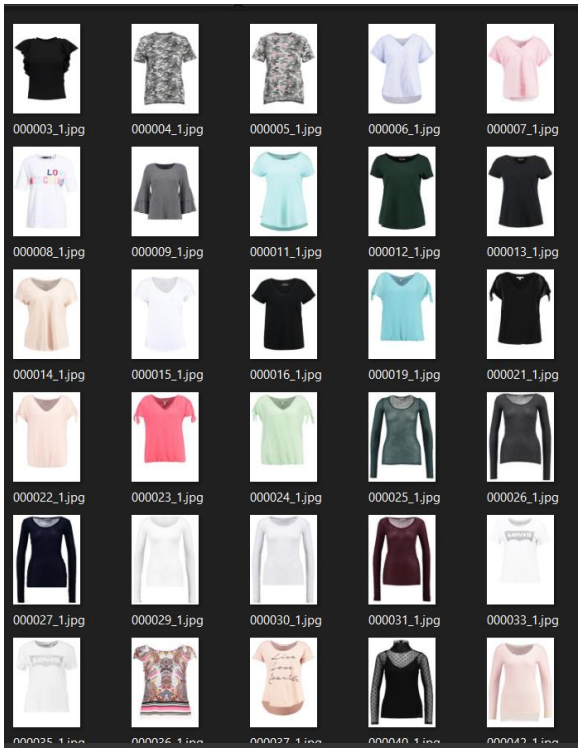


Figure 3.2.1.1: Training image of the cloth

Figure 3.2.1.2: Training image of the cloth mask

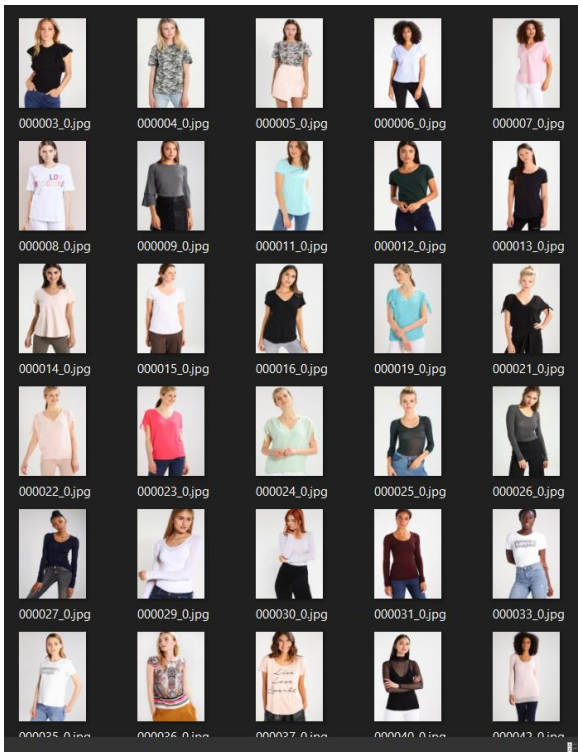


Figure 3.2.1.3: Training image of the person wearing the cloth.

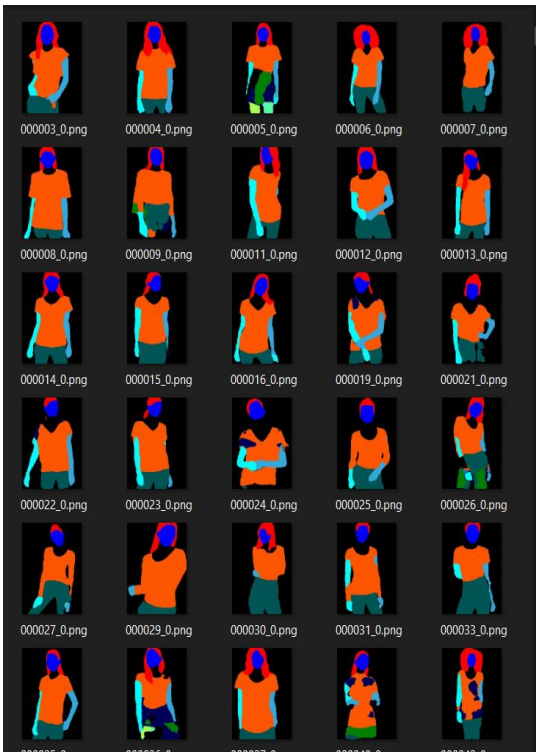


Figure 3.2.1.4: Training image of the person parse

3.2.2 Person Representation

Person representations contain a set of features including pose heatmap, human body representation and face and hair segment. (Han et al., 2018)

Pose heatmap. the pose heat is a pose estimator. The pose of the person is show using the coordinates of 18 key point. The key point is transformed into a heatmap to leverage the spatial layout. Lastly a 18-channel pose heatmap is produce by stacking the heatmaps from all key point.

Human body representation. A human segmentation map is producing by using the human parser. Different parts of human body like arms, legs and so on are represented by the different regions. Next, a 1-channel binary mask is producing by convert the segmentation map. one in the mask is indicate human body while zeros are elsewhere. In order to avoid the artifacts when the body shape and target clothing conflict, the binary mask is directly down sampled to a lower resolution.

Face and hair segment. The face and hair segment is use to maintain the identity of the person. Physical attributes like face, skin color, hair style and others are incorporated into the system.

The RGB channels of face and hair regions of a person is extracted by the system with the use of the human parser. The RGB channels is use because is to inject identity information when generating new images.

Lastly, the system will resize these three feature maps. It will resize them into a same resolution and it also will concatenate them to form a clothing- agnostic person representation.

3.2.3 Geometric Matching Module (GMM)

Align clothing while preserving the characteristic is the main purpose of the Geometric Matching Module. A convolutional neural network is adopted by the system so that it able to

Figure 3.2.3.1: The input grid and clothing images are warped stably with the proposed GIC loss, whereas a swirling pattern is presented in the warped grid without GIC loss. (Lee et al., 2019)

learn the transformation parameters, including feature extracting layers, feature matching layers, and the transformation parameters estimating layers. The system takes the person representation and in-shop clothing as input and output the warped clothes. So, first person representation and in-shop clothing passed through the feature extracting layers. After that, the system will use matching layer to predicts the correlations map. Lastly, the system will guess the Thin-Plate Spline (TPS) transformation parameters for the clothes picture directly by using a regression network. TPS transformation generally shows high performance but its high flexibility often causes distortion of patterns and prints. Thus, to retain the characteristics of clothes after warping, grid interval consistency (GIC) loss is introduced (Lee et al., 2019).

Other than that, occlusion handing is introduce as well, this is because the cloths are very simply occluded by arms and hair. The occluded regions are excluded from the warp loss calculation. This will produce accurate geometric transformation no matter the occlusion by hair and arms.

Figure 3.2.3.2: Occluded clothes on person cause unreasonable transformation and distortion. With occlusion handling method, the network conducts accurate geometric matching. (Lee et al., 2019)

3.2.4 Try-on Module (TOM)

The main point of doing TOM is to combine the warp cloth form the GMM to the person image. First of all, in order to preserve the other human features other than the target clothing area like face, lower cloths, hair, and legs are add on to human representation input of TOM. TOM use U-net to avoid unnatural seams appear instead directly copied onto the target person-image. After the U-net the composition mask will be produce and the composition mask is combining with warped clothing and intermediate person-image to produce a seamless image. To improve the quality of image generated a synthesizing discriminator is use at the end of TOM. The synthesizing discriminator is a SN-PatchGAN which take in the result and ground truth image.

3.3 IMPLEMENTATION ISSUES AND CHALLENGES

- Unable to deform the input clothing to the 3D pose of the target person. To manipulate clothing and human shape and detail are an extremely hard task, because of the huge variety of pose of target person. Our system suffers from varieties of wardrobe and person style especially 3D poses like crossing both arms. This will cause the system to produce an image that will cover the hand. The system will produce distorted image when the person image has the person crossing their arm around their body like in figure

person image	Cloth image	Final image
		

Figure 3.3.1: Distorted image

3.4 TIMELINE

This is an overview of the project schedule for the preparation to develop the virtual try on system within a duration of 14 weeks. Each of the task is expected to be completed within the period as stated in the project schedule below.

Table 3.4.1: Project Timeline

Project Task	**Project Week**													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Planning														
Forming project title														
Research with existing method														
Discussion with Supervisor														
Analysis														
Identify project background														
Identify problem statement														
Literature review of existing system														

Creating system functionality comparison table													
Identify project objective													
Identify project scope													
Identify project innovation and contribution													
Critical remark of previous works													
Identify flow chart													
Identify enhanced flow chart													
Implementation													
Coding the system													
Perform system Testing													
Deploy the system													
Prepare Final Year Report													
Submit Final Year Report													

CHAPTER 4: SYSTEM IMPLEMENTATION, TESTING AND RESULT

4.1 System Implementation

Install the python library and required software

This system is developed using Anaconda for the environment and python programming language.

Install Anaconda

The used of Anaconda is for environment management system that installs, runs, and updates packages.

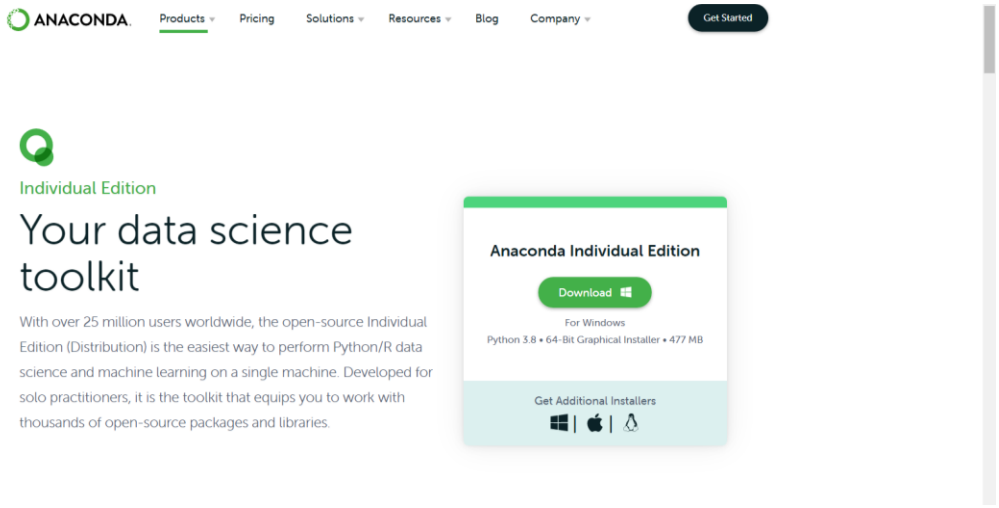


Figure 4.1.1: Anaconda Download Website

Install Library Packages

The library packages needed to the system are PyTorch Library, CuPy Library, OpenCV Library, Scikit-image Library, Scipy Library, Pillow Library, Mediapipe Library, and Flask Library. These libraries can be download after a new conda environment is created in the anaconda terminal.

Step 1: conda create -n tryon python=3.7

Step 2: conda activate tryon

Step 3: conda install pip

Step 4: follow table 4.1.2

Table 4.1.2: commands to download the library packages

Library Packages	Command
PyTorch	conda install pytorch=1.1.0 torchvision=0.3.0 cudatoolkit=9.0 -c pytorch
CuPy	pip install cupy==6.0.0
OpenCV	pip install opencv-python==4.5.1
Scikit-image	pip install scikit-image==0.18.1
Scipy	pip install scipy==1.6.2
Pillow	pip install pillow==8.3.1
Mediapipe	pip install mediapipe==0.8.7
Flask	pip install flask==2.0.1

Download the system

User can get the code from github:

Pip clone https://github.com/Oharahaoyonggan/tryon-cloth.git

cd tryon-cloth

After that download the pretrain checkpoint and unzip the checkpoint and paste it in the checkpoint folder in tryon-cloth

https://drive.google.com/file/d/1JGxhpblMuxYZefyxLbgd1PksPkOBfw89/view?usp=sharing

Running the System in flask framework

The system is run on flask because is the best web framework. Flask is written in Python and it easy to develop web applications easily. This is because flask is a microframework that doesn't include object relational manager. To run the system in the flask framework, user need to run the flask command on the terminal with the anaconda environment in the virtual tryon directory.

Command to run flask: flask run

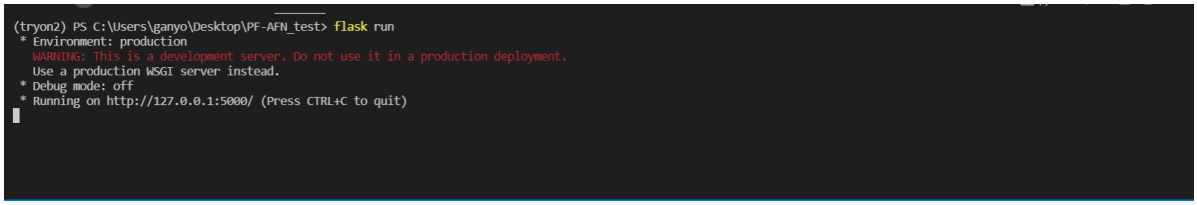


Figure 4.1.3: Running flask

Web application for user to upload photo

User able to upload the person image (their own image) and cloth image (cloth they wanted to wear)

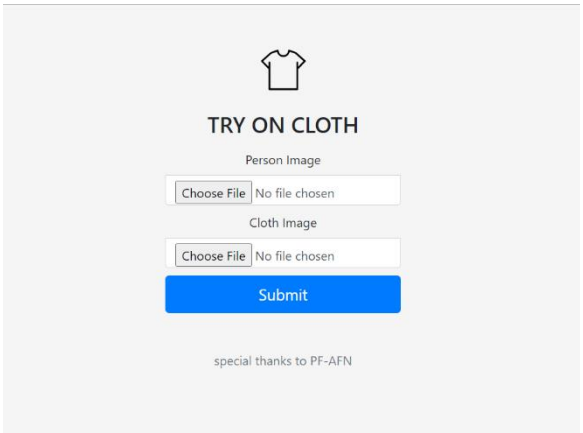


Figure 4.1.4: Virtual try on web app

Example of person image:

Figure 4.1.5: Person image

Example of cloth image:

Figure 4.1.6: Cloth image

Result will be display in the web app and a go back to return back to main page

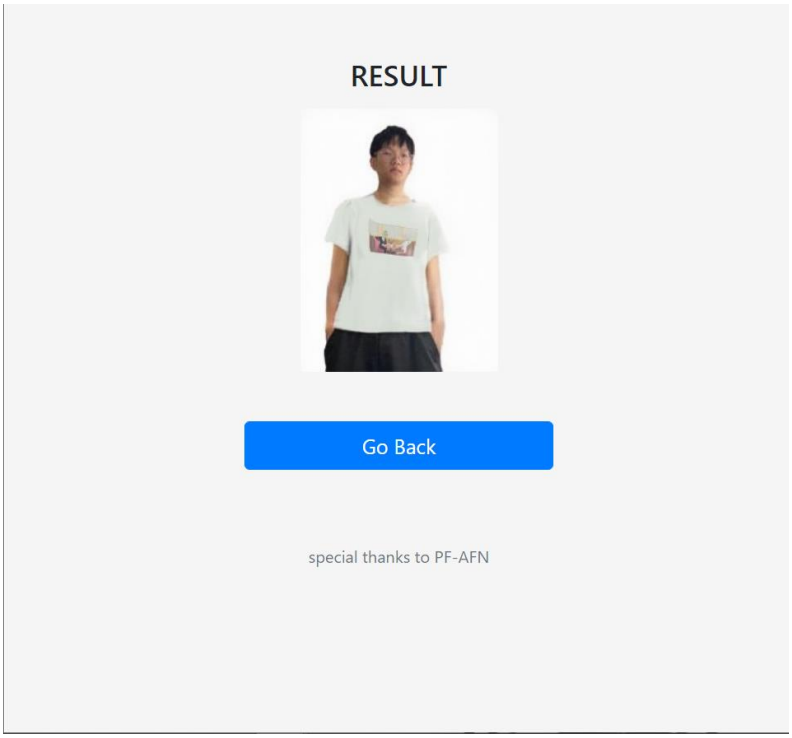


Figure 4.1.7: Try-on result

4.2 Testing

For testing, there are total 2032 pairs of images, every pair had a top clothing image with 256 X 192 resolution and a front-view women photo. A lot of previous virtual try-on system use this dataset to test their system. The target cloths image and the reference person image are given as the input of the system to generate the image.

Test case 1 to test the handle cloth image with complex pattern and the final output image should be able to preserve complex pattern of the cloth.

test case 1:

Table 4.2.1: Test case 1 – cloth image with complex pattern

Person image	Cloth image	Final image
		
		

Based on test case it shows that the system able to preserve the complex pattern of the cloth. Like in table 4.2.2 last row the texture of the floral shirt is able to be preserve at the final image.

Test case 2 to test the occlusion handling where it able to produce a precise geometric transformation nor matter of occlusion by arms and hair.

test case 2:

Table 4.2.2: Test case 2 – Person image with complex pose

Person image	Cloth image	Final image
		

Based on test case 2 it shows that the system able to handle some of the occlusion like in table 4.2.2 row one the system able to handle simple hand occlusion as it does not block the body too much, while for row two the system unable to do the hand occlusion as the hand blocking a big part of the body resulting distorted image. The last row show that the system able to handle the hair occlusion.

Test case 3 is to test the preserve of other human features other than the target clothing area like face, hair, lower cloths, and legs.

test case 3:

Table 4.2.3: Test case 3 – Person image with other human features

Person image	Cloth image	Final image
		
		

Based on test case 3 it shows that the system able to preserve of other human features other than the target clothing area like face, hair, lower cloths, and legs. In table 4.2.3 the last image show that the system able to preserve the face, hand and lower cloths.

4.3 Results

Qualitative Results

The system proposed generate a highly realistic try-on result. At the same time the system able to handle large misalignment between the person and the cloth, preserves the characteristics of both the non-target cloth like hair, face and lower body part and the target cloths while retaining the clear body parts. Other than that, it also able to model long-range correspondence between the person and the cloths, this enables the system to avoid the distortion in embroideries and logo. Based on figure 4.3.1, the characteristic of the cloth is preserved without any distortion. Also, the human feature such as face, lower clothing and hand is able to be preserve as well.

Success case:

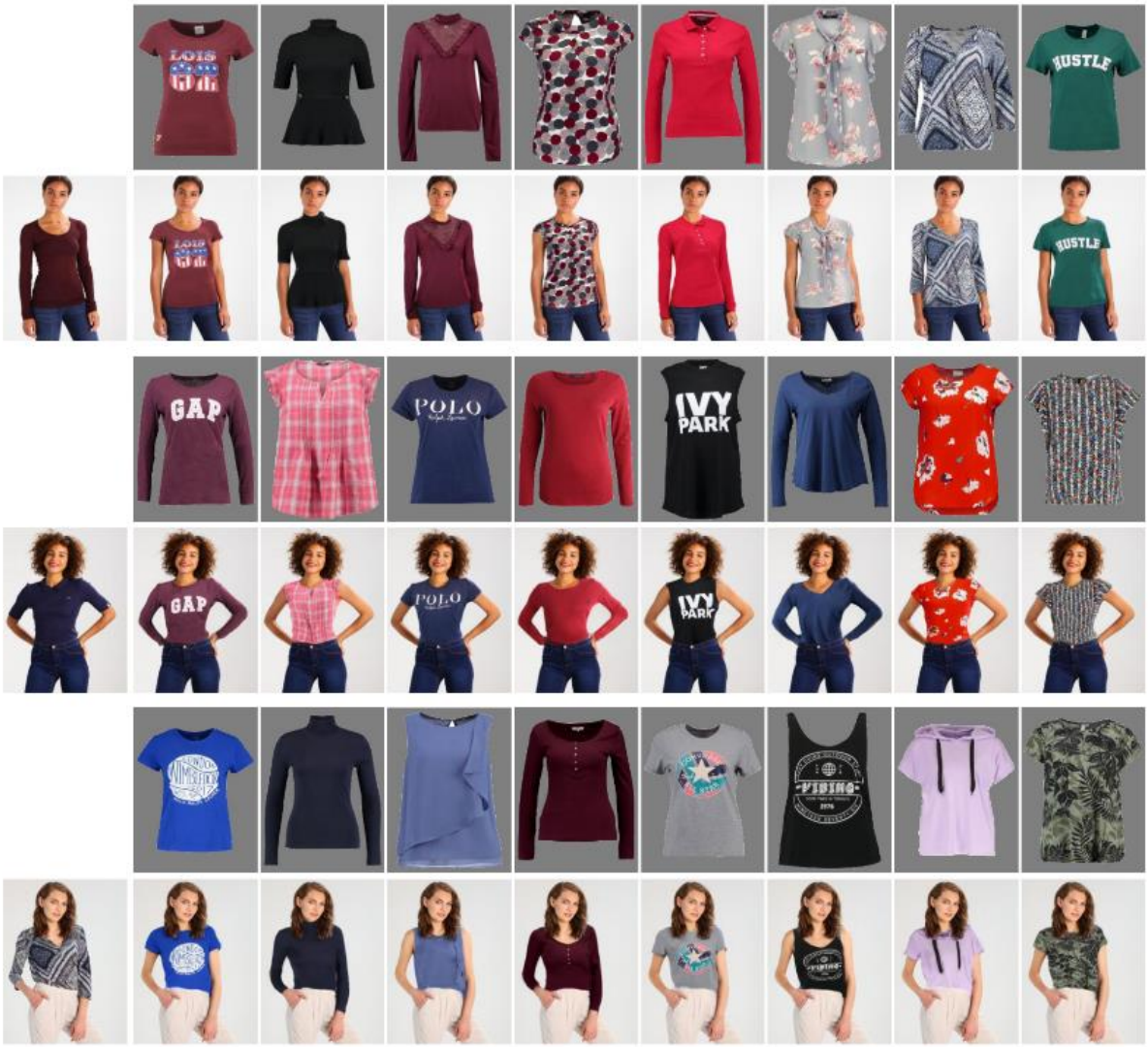


Figure 4.3.1: success case of the system

But the system unable to completely solve the occlusion issue as there is still some defect especially when there is an arm in front of the cloth like figure 4.3.2.

Not successful case:

Figure 4.3.2: Not successful case

Quantitative Analysis

For a 2D image based virtual try-on, a reference person image and a target cloth are used as the input to generate the try-on results during the test. Because of the ground truth image which is reference person wearing the target cloths is unavailable. Thus, the FID is used to evaluate the metric. The lower the score of FID, mean the higher the quality of the results. Based on the FID our system has the lowest FID. Base on table 4.3.3 the FID score of our system has the lower score compare to other system like CP-VTON and CP-VTON+ which mean the image produce by our system is very similar to the original image.

Method	FID
CP-VTON	24.43
CP-VTON+	21.08
Our system	10.09

Table 4.3.3: FID score

CHAPTER 5: CONCLUSION

5.1 Conclusion

There are more and more brick-and-mortar retail shops had moved their business online because of the pandemic. Comparing the e-commerce sector, online apparel has the biggest share and it also continue going up with tremendous pace. This is because online apparel shopping able to provide the ease of shopping from the comfort of one's home. But the downside of it is that customer is limited to understand how a cloth will look like on them because the online shopping does not allow trying the cloth on physically.

The motivation of this system is that there are more and more useful cloths datasets for training like DeepFashion (Ziwei Liu, 2016). With this it allows the possibility of doing virtual try-on task with only using 2D photo which is cheaper. This is because traditional virtual try-on system relies on 3D information of the cloth and the body shape. The 3D information is expensive and take time as the body shape and garment needed to be scanned to produce as 3D model. Other than that, the current virtual try-on method fails to preserve the small detail like the hair of the person lose the details and style and the clothing of lower-body. Most of the current method, use a coarse-to-fine network to produce the picture to generate a realistic image. The current system totally ditches the important feature of human parsing, which will produce a unreasonable and blurry image.

To solve the problem the system proposed able to learns to synthesize the new person image from virtual try-on by manipulating cloths, given an input person image and a desired cloth, the proposed system aims to produce a new image of the person wearing the desired clothes. The system that we propose is inspired by CP-VTON+ (Matiur et al., 2020) and LA-VTON (Lee et al., 2019). The system is separate into two part the first part is Geometric Matching Module (GMM) and the second part is Try-on module (TOM). For GMM, the main use of this module is to align the desired cloth to the person body and output a warped cloth. While for the TOM, the main use of this module is to use the warped cloth produce by GMM and person representation, then combine it to produce the person wearing the desired cloth. Other than this there are a few techniques are added to the system to make the image produce more perfect. One of it is grid interval consistency is added to GMM to retain the characteristics of the warp cloth produce by GMM. Not only that, Occlusion handling is also added to GMM as it able to help in cloths that is occluded by hair and arms. So, the warp cloth produce by GMM won't

deform to much because of the clothing area in the cloths segmentation of person image is being block by hair or arms. Lastly, the system also has added a synthesizing discriminator at the end of TOM using SN-PatchGAN to further improve the quality of images generated.

The advantage of the system is it able to keep the feature of the clothing product for example the logo, text, and texture without noticeable artefacts and distortion. The system also able to preserve other human features other than the target clothing area like the human face, lower cloths, hair, and legs. Aside from that, the system also able to do some occlusion handling like occlusion handling of the hair. But the downside of is unable to change the input shirt to the 3D pose of the target person. If the input person image has the pose where most of the body is been block by the hand, the system will not able to do the occlusion and produce distorted result.

5.2 Future Work

One of the problems of the proposed system is unable to change the input shirt to the 3D pose of the target person. To manipulate clothing shape and texture and human are a super challenging thing to do, because of the huge variety of pose of the target person. Our system suffers from different style of cloth and human pose especially 3D poses like crossing both arms. This will cause the system to produce an image that will cover the hand. For future enhancement of this system, we can explore the use of 3D structure of the reference person to handle occlusion in virtual try-on. Other than that, the current system unable to generate layer cloth such as jacket with t-shirt below it. Therefore, the use of 3D structure might solve this issue as well.

REFERENCES

Han, X., Wu, Z., Wu, Z., Yu, R., & Larry S. (2018) Viton: An image-based virtual try-on network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7543–7552.

Wang, B., Zheng, H., Liang, X., Chen, Y., Lin, L., & Yang, M. (2018) Toward characteristic preserving image-based virtual try-on network. In Proceedings of the European Conference on Computer Vision (ECCV), pp.589–604.

Thibaut, I., Jer´emie, M., & Cl´ement, C. (2019) End-to-end learning of geometric deformations of feature maps for virtual try-on. arXiv preprint arXiv:1906.01347.

Dong, H., Liang, X., Wang, B., Lai, H., Zhu, J., & Yin, J. (2019) Towards multi-pose guided virtual try-on network. arXiv preprint arXiv:1902.11026.

Masahiro, S, Kaoru, S., Frank, P., Bjorn, S., & Masashi, N. (2014) Virtual fitting by single-shot body shape estimation. In Int. Conf. on 3D Body Scanning Technologies, pp 406–413.

Lee, H., Lee, R., Kang, M., Cho, M., & Park, G.(*2019*) LA-VITON: A Network for Looking-Attractive Virtual Try-On, *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp.3129-3132.

Wang, J., Zhang, W., Liu, W., & Mei, T. (2019) Down to the Last Detail: Virtual Try-on with Detail Carving. *ArXiv, abs/1912.06324*.

Z. Liu, P. Luo, S. Qiu, X. Wang, & X. Tang , (2016) Deepfashion: Powering robust clothes recognition and retrieval with rich annotations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1096–1104.

Rocco, R. Arandjelović and J. Sivic (2017) Convolutional neural network architecture for geometric matching In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Matiur, R. Minar, Thai, T. Tuan, H. Ahn, Paul, R., & Y. Lai. (2020) "CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On." In: CVPR Workshop on Computer Vision for Fashion, Art, and Design (CVPRW).

Ronneberger, O., Fischer, P. and Brox, T., (2015) U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241).

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P., (2004) Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), pp.600-612.

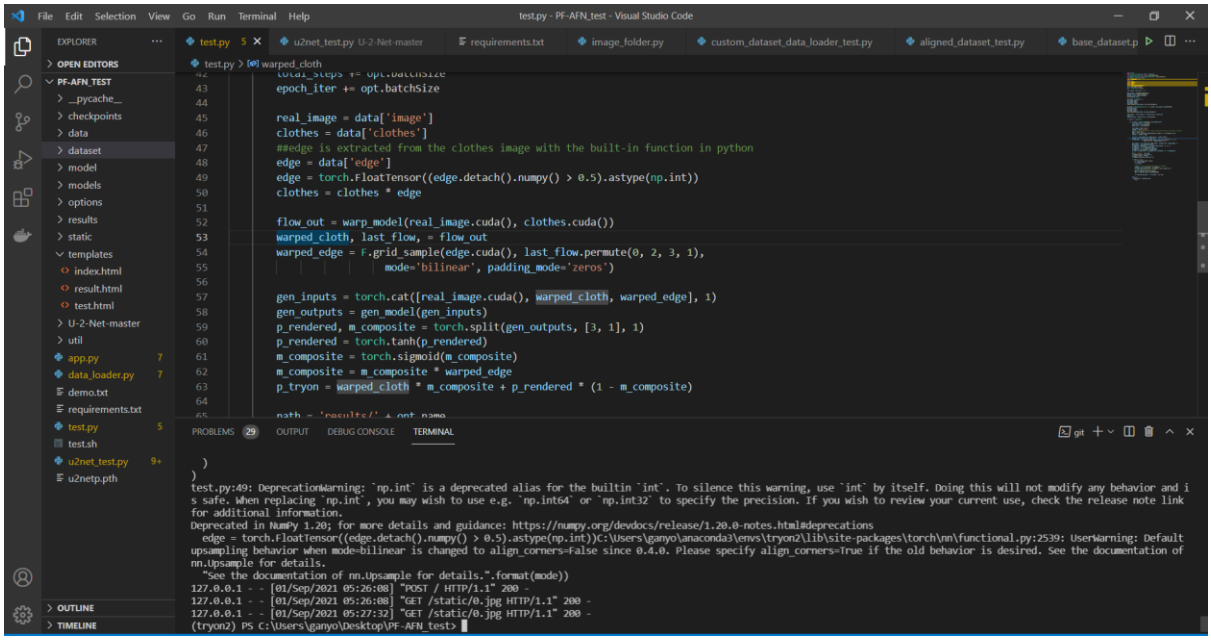
Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X., (2016). Improved techniques for training gans. arXiv preprint arXiv:1606.03498.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S., (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. arXiv preprint arXiv:1706.08500.

Schallom, R., 2020. Many businesses turn to e-commerce for the first time due to the pandemic. [online] Fortune. Available at: <https://fortune.com/2020/07/15/ecommerce-online-shopping-coronavirus-business-trends-covid/> [Accessed 7 April 2021].

APPENDICES A

Using Visual Studio code



FINAL YEAR PROJECT WEEKLY REPORT

(Project I,II)

Trimester, Year: S2, Y3	**Study week no.: 3**
Student Name & ID: GAN YONG HAO ,18ACB00330	
Supervisor: Ts Lai Siew Cheng	
Project Title: Image-Based Virtual Try-On System using Deep Learning.	

1. WORK DONE

Finalize the TOM of the system

2. WORK TO BE DONE

tune the system so that it able to do occlusion handling and GIC

3. PROBLEMS ENCOUNTERED

none

4. SELF EVALUATION OF THE PROGRESS

Going smoothly as plan

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project I,II)

Trimester, Year: S2, Y3	**Study week no.: 6**
Student Name & ID: GAN YONG HAO ,18ACB00330	
Supervisor: Ts Lai Siew Cheng	
Project Title: Image-Based Virtual Try-On System using Deep Learning.	

1. WORK DONE

Tuning the wrap using the GIC loss and done some occlusion handling

2. WORK TO BE DONE

Testing the system

3. PROBLEMS ENCOUNTERED

none

4. SELF EVALUATION OF THE PROGRESS

Going smoothly as plan

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project I,II)

Trimester, Year: S2, Y3	**Study week no.: 9**
Student Name & ID: GAN YONG HAO ,18ACB00330	
Supervisor: Ts Lai Siew Cheng	
Project Title: Image-Based Virtual Try-On System using Deep Learning.	

1. WORK DONE

Testing the system and find the FID score

2. WORK TO BE DONE

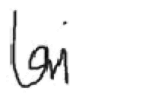
Implement the system using web app

3. PROBLEMS ENCOUNTERED

none

4. SELF EVALUATION OF THE PROGRESS

Going smoothly as plan



_____

_Supervisor's signature

_____

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project I,II)

Trimester, Year: S2, Y3	**Study week no.: 11**
Student Name & ID: GAN YONG HAO ,18ACB00330	
Supervisor: Ts Lai Siew Cheng	
Project Title: Image-Based Virtual Try-On System using Deep Learning.	

1. WORK DONE

The system is implemented as a web app using flask

2. WORK TO BE DONE

Start writing the fyp2 report

3. PROBLEMS ENCOUNTERED

none

4. SELF EVALUATION OF THE PROGRESS

Going smoothly as plan

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project I,II)

Trimester, Year: S2, Y3	**Study week no.: 13**
Student Name & ID: GAN YONG HAO ,18ACB00330	
Supervisor: Ts Lai Siew Cheng	
Project Title: Image-Based Virtual Try-On System using Deep Learning.	

1. WORK DONE

Finished writing the report for FYP2

2. WORK TO BE DONE

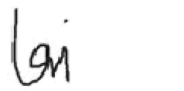
Prepare for the presentation

3. PROBLEMS ENCOUNTERED

none

4. SELF EVALUATION OF THE PROGRESS

Going smoothly as plan



Supervisor's signature

Student's signature

APPENDIX C POSTER

Tired of guessing whether the cloth look on you when shopping online!

IMAGE-BASED VIRTUAL TRY-ON SYSTEM USING DEEP LEARNING

- You able to transfer the in-shop clothing image to the your photo.
- The system also able to do occlusion handling and retain the in-shop cloth feature.
- The system able to preserve you image, so you won't look distorted

Example:

APPENDIX D PLAGIARISM CHECK RESULT

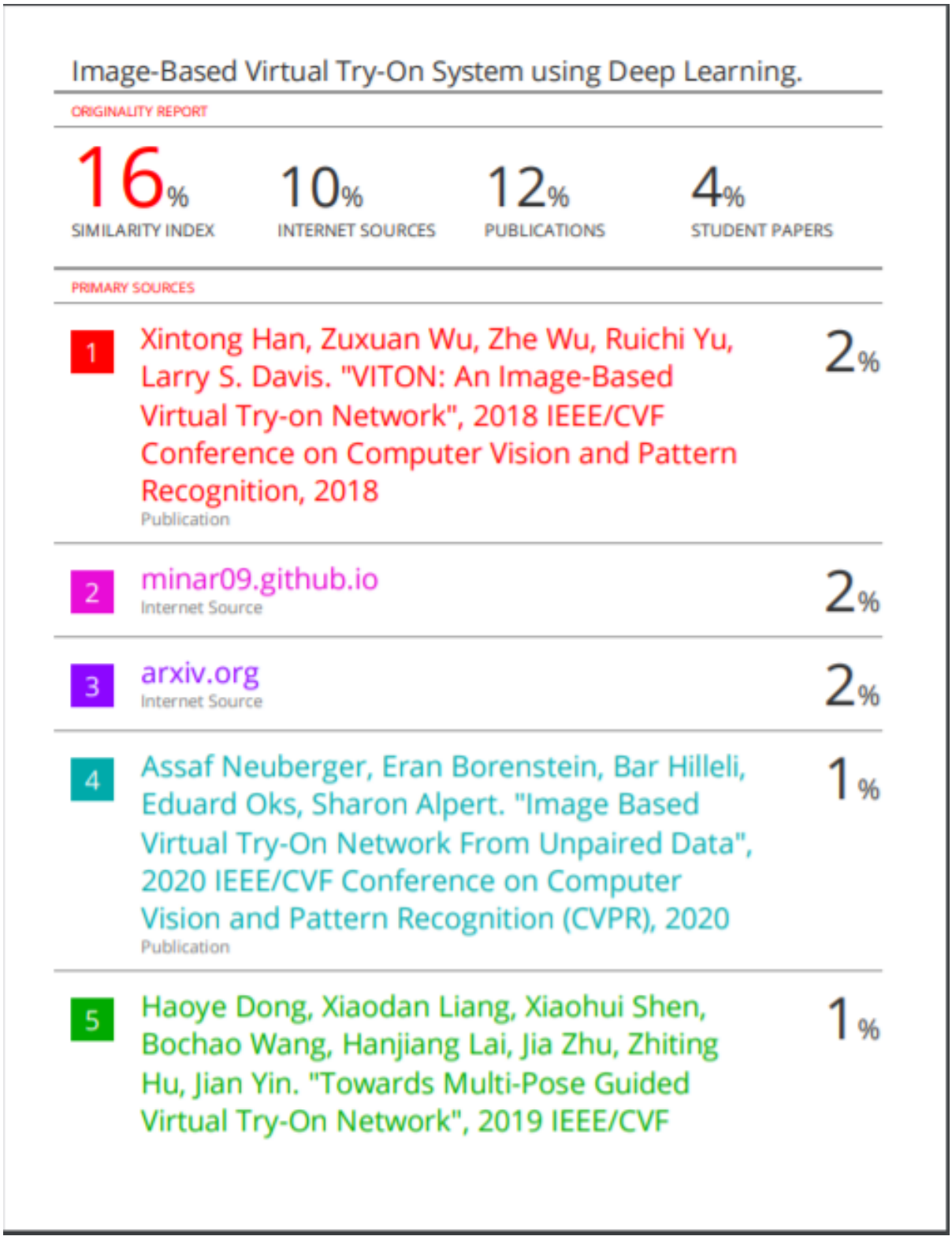


Image-Based Virtual Try-On System using Deep Learning.

ORIGINALITY REPORT

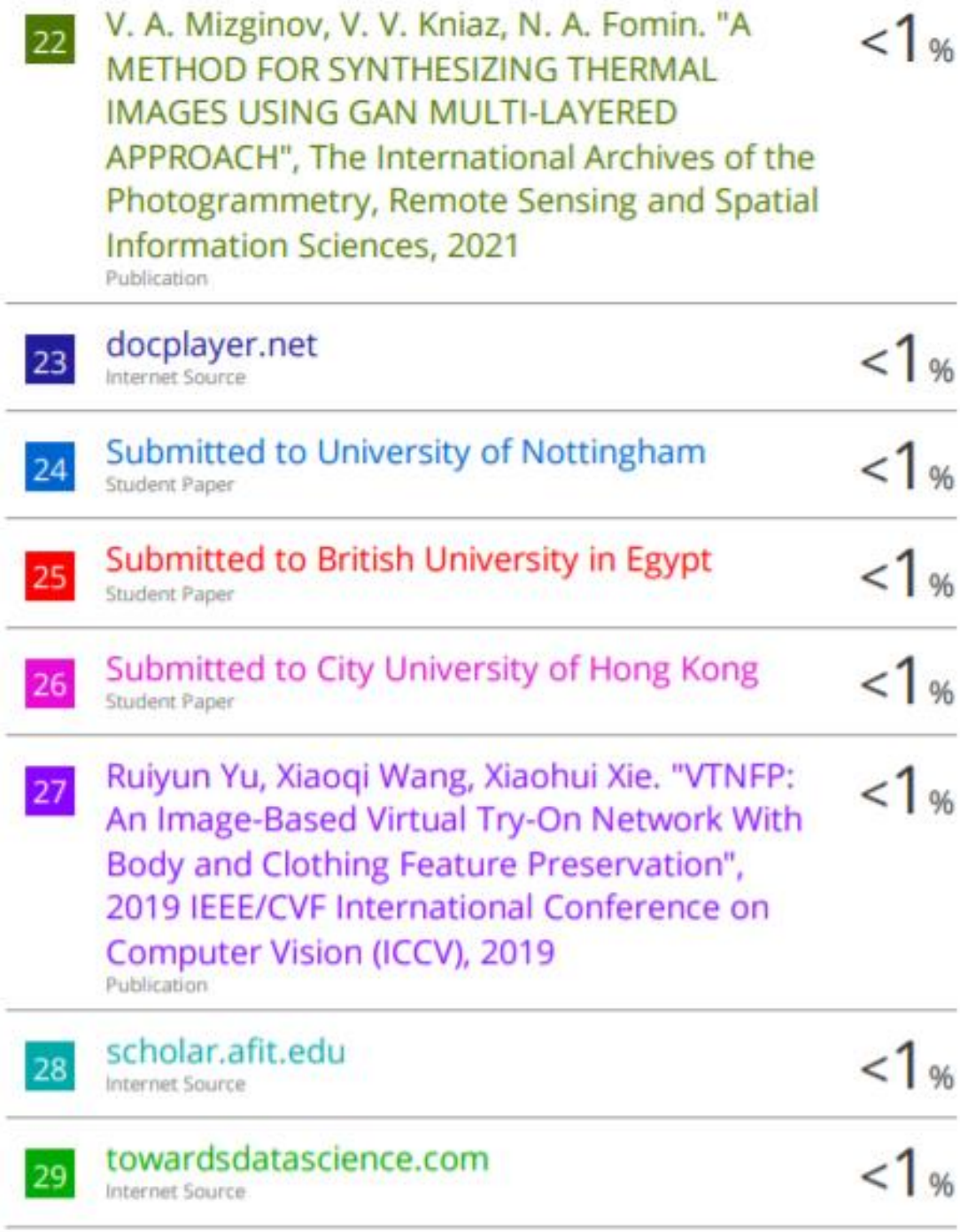
16%	10%	12%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

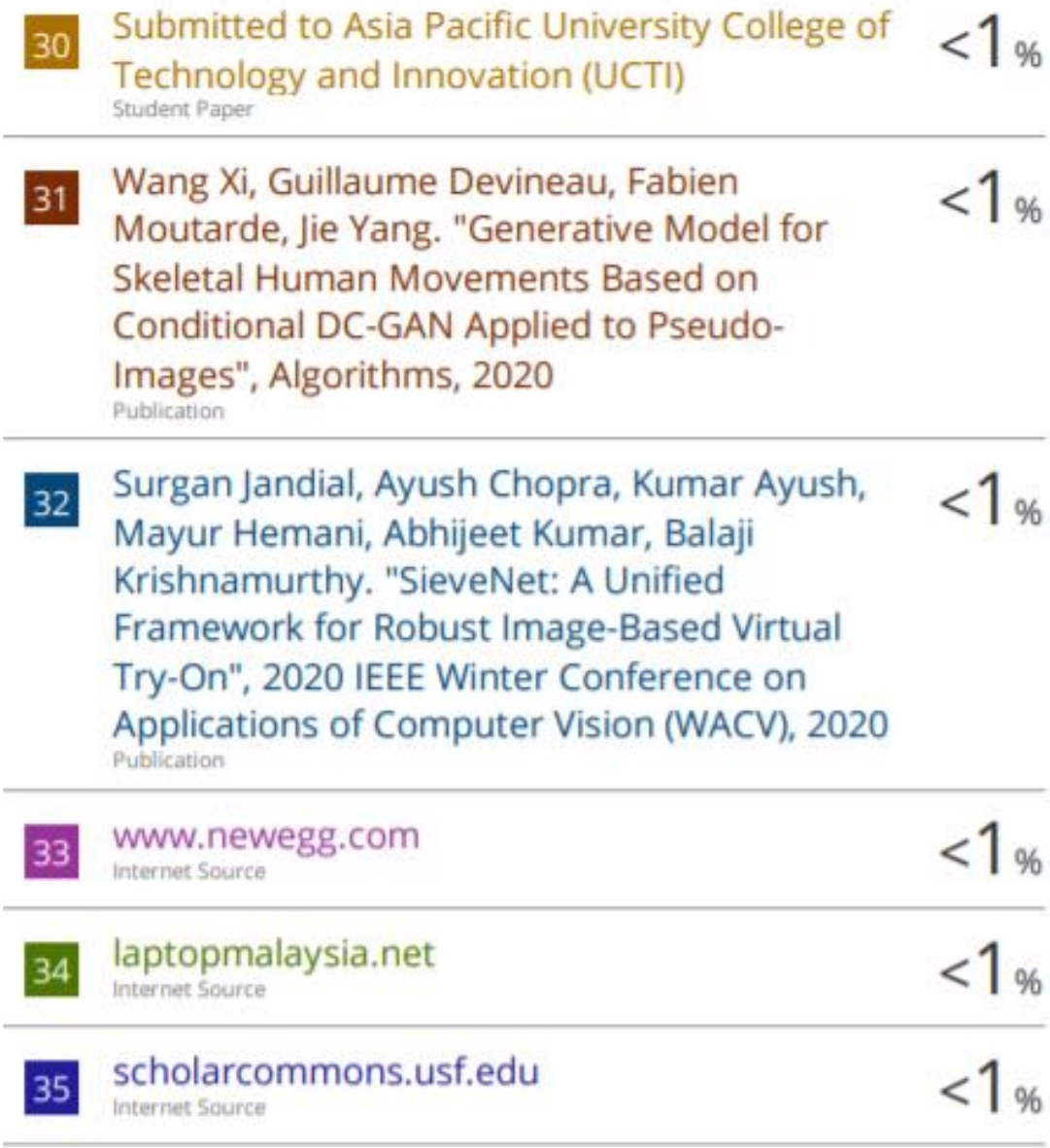
1	Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, Larry S. Davis. "VITON: An Image-Based Virtual Try-on Network", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018 Publication	2%
2	minar09.github.io Internet Source	2%
3	arxiv.org Internet Source	2%
4	Assaf Neuberger, Eran Borenstein, Bar Hilleli, Eduard Oks, Sharon Alpert. "Image Based Virtual Try-On Network From Unpaired Data", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020 Publication	1%
5	Haoye Dong, Xiaodan Liang, Xiaohui Shen, Bochao Wang, Hanjiang Lai, Jia Zhu, Zhiting Hu, Jian Yin. "Towards Multi-Pose Guided Virtual Try-On Network", 2019 IEEE/CVF	1%


International Conference on Computer Vision (ICCV), 2019
Publication
6 eprints.utar.edu.my
Internet Source
1%
7 Submitted to Rochester Institute of Technology
Student Paper
1%
8 medium.com
Internet Source
<1%
9 Hyug Jae Lee, Rokkyu Lee, Minseok Kang, Myounghoon Cho, Gunhan Park. "LA-VITON: A Network for Looking-Attractive Virtual Try-On", 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019
Publication
<1%
10 deepai.org
Internet Source
<1%
11 Kumar Ayush, Surgan Jandial, Ayush Chopra, Balaji Krishnamurthy. "Powering Virtual Try-On via Auxiliary Human Segmentation Learning", 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019
Publication
<1%
12 Submitted to Universiti Tunku Abdul Rahman
Student Paper
<1%

13	www.groundai.com Internet Source	<1 %
14	openaccess.thecvf.com Internet Source	<1 %
15	github.com Internet Source	<1 %
16	"Pattern Recognition and Computer Vision", Springer Science and Business Media LLC, 2020 Publication	<1 %
17	fortune.com Internet Source	<1 %
18	link.springer.com Internet Source	<1 %
19	publications.naturalengland.org.uk Internet Source	<1 %
20	en.wikipedia.org Internet Source	<1 %
21	Na Zheng, Xuemeng Song, Zhaozheng Chen, Linmei Hu, Da Cao, Liqiang Nie. "Virtually Trying on New Clothing with Arbitrary Poses", Proceedings of the 27th ACM International Conference on Multimedia - MM '19, 2019 Publication	<1 %



22	V. A. Mizginov, V. V. Kniaz, N. A. Fomin. "A METHOD FOR SYNTHESIZING THERMAL IMAGES USING GAN MULTI-LAYERED APPROACH", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2021 Publication	<1 %
23	docplayer.net Internet Source	<1 %
24	Submitted to University of Nottingham Student Paper	<1 %
25	Submitted to British University in Egypt Student Paper	<1 %
26	Submitted to City University of Hong Kong Student Paper	<1 %
27	Ruiyun Yu, Xiaoqi Wang, Xiaohui Xie. "VTNFP: An Image-Based Virtual Try-On Network With Body and Clothing Feature Preservation", 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019 Publication	<1 %
28	scholar.afit.edu Internet Source	<1 %
29	towardsdatascience.com Internet Source	<1 %



30	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
31	Wang Xi, Guillaume Devineau, Fabien Moutarde, Jie Yang. "Generative Model for Skeletal Human Movements Based on Conditional DC-GAN Applied to Pseudo-Images", Algorithms, 2020 Publication	<1 %
32	Surgan Jandial, Ayush Chopra, Kumar Ayush, Mayur Hemani, Abhijeet Kumar, Balaji Krishnamurthy. "SieveNet: A Unified Framework for Robust Image-Based Virtual Try-On", 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 2020 Publication	<1 %
33	www.newegg.com Internet Source	<1 %
34	laptopmalaysia.net Internet Source	<1 %
35	scholarcommons.usf.edu Internet Source	<1 %

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	GAN YONG HAO
ID Number(s)	18ACB00330
Programme / Course	FICT Bachelor of Computer Science (Honours) (CS)
Title of Final Year Project	IMAGE-BASED VIRTUAL TRY-ON SYSTEM USING DEEP LEARNING

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 16 % **Similarity by source** Internet Sources: 10 % Publications: 12 % Student Papers: 4 %	
Number of individual sources listed of more than 3% similarity: 0	
Parameters of originality required and limits approved by UTAR are as Follows: **(iii) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Lai Siew Cheng

Date: 3/9/2021

Signature of Co-Supervisor

Name: _____

Date: _____

M

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	18ACB00330
Student Name	GAN YONG HAO
Supervisor Name	Ts Lai Siew Cheng

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
	Front Plastic Cover (for hardcopy)
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
√	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report. (Signature of Student: GAN YONG HAO) Date:02-09-2021	Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction. (Signature of Supervisor) Date: 3/9/2021