Department of Computer Science

MSc [Artificial Intelligence]

Academic Year 2024-2025

[Virtual try-on models vs e-commerce]

[Amil Kazimoglu 2449223]

A report submitted in partial fulfilment of the requirement for the degree of Master of Science

Brunel University
Department of Computer Science
Uxbridge, Middlesex UB8 3PH
United Kingdom
Tel: +44 (0) 1895 203397
Fax: +44 (0) 1895 251686

ABSTRACT

The growth of e-commerce has made a demand for technologies which are improving the customer confidence and reducing the return rates. Virtual try-on technology could be a one solution to that., allowing users to visualise the product before paying it. However, most existing studies have focused on mainly image quality metrics, while overlooking critical factors such as scalability, inference speed and real-world deployment.

This research tries to evaluate these existing virtual try on models in a different category wise, such as GAN based, diffusion based and prompt-based model, along with them LoRA fine-tuned variant of prompt-based model has evaluated as well. Quantitative and qualitative evaluations were conducted, supported by performance and scalibity analyses. In addition to that product of concept has been developed, which is web application where all models are embedded to demonstrate the user insights and highlights the trade-offs quality, performance and scalability results by making them visual in e-commerce web application. Also, LoRA fine tuning has been done to not just using and evaluating ready models but doing a contribution to create a new model to evaluate experiments and trying to improve them by this approach.

The findings show that no single model is universally optimal model, GANs are fast and lightweight but produce weaker realism, diffusion models achieve higher fidelity at greater computational cost, and prompt-based models add flexibility but remain resource-intensive. LoRA fine-tuning yielded small improvements, suggesting potential for targeted adaptation under constraints. Also, further improvements are needed to evaluate the quality metrics that has been use in the literature, from the qualitative findings and comparing with quantitative ones, this shows the limitation of quality metrics are not very trustable when evaluating the human perception.

The study concludes that hybrid deployment strategies which is using all models in web application and give them to user by business idea, such as using fast GAN models as free previews and using diffusion and prompt-based models as premium to use in e-commerce.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, YongminLli for his continuous support, feedback and invaluable guidance throughout this course. Their expertise and encouragement have been instrumental in shaping the direction and quality of this work.

I am also grateful to the academic staff and fellow students in the MSc Artificial Intelligence programme at Brunel University London for creating an intellectually stimulating environment that has inspired me to challenge myself and grow both academically and personally.

Sign in the box below to certify that the work carried out is your own. By signing this box you are certifying that your dissertation is free from plagiarism. Make sure that you are fully aware of the Department guidelines on plagiarism (see the student handbook). The penalties if you are caught are severe. All material from other sources must be properly referenced and direct quotes must appear in quotation marks.

I certify that the work presented in the dissertation is my own unless referenced

Signature....Amil Kazimoglu.....

Date.....25/09/2025.....

Insert a word count. This is the sum of the words in all the chapters only. The sum should exclude the words in the title page, abstract, acknowledgements, table of contents, references and any appendices.

TOTAL NUMBER OF WORDS:11005

[*All the above should not exceed this one page]*

TABLE OF CONTENTS

APPENDIX A: ETHICAL APPROVAL

Please see the Ethics website for details of what should appear here.

APPENDIX B: etc.

The rest of your appendices should contain details such as questionnaires, results from interviews, documented source code, documented design, etc. where applicable.

CHAPTER 1: INTRODUCTION

The fast growth of the digital world has affected the retail industry by reshaping customer expectations and industry practices regarding product interactions. From this evolution in the fashion sector, new technologies have emerged to satisfy the dynamic demands of the field, where visualisation tools are becoming central to customer engagement and their purchase decisions. Basegmez and Tuncali Yaman (2022) support the idea that visualisation tools influence purchasing decisions on e-commerce platforms, indicating that one of the technologies widely adopted on such platforms is virtual try-on (VTON) models.

Virtual try-on is a technology that allows users to visualise garments on digital avatars or images of themselves. Islam et al. (2024b) refer to the potential of this technology not only to affect consumers' purchasing decisions but also to provide a means for consumers to engage with products virtually, thereby strengthening the connection between businesses and customers. Islam et al. (2024b) also note that VTON systems have been developed using multiple technical approaches, each attempting to improve upon earlier methods. Although these models have been continuously improved in the literature, they are often evaluated primarily in terms of technical image quality metrics. What is largely missing is an evaluation of their real-world suitability for e-commerce.

This gap is particularly important because Nandhakumar et al. (2025) indicate that virtual try-on models are increasingly being integrated into e-commerce platforms. They emphasise that the use of VTON systems enables customers to make instant decisions about how a garment will appear on them, without speculation. However, without evaluating these systems in e-commerce platforms using additional metrics such as inference time, GPU memory consumption, model size, and responsiveness, customers' decisions are left uncertain. Despite this, much of the literature continues to focus on model-to-model competition rather than evaluating the broader commercial potential of VTON systems.

To address this gap, this dissertation adopts a dual perspective: evaluating both the visual quality and practical performance of representative VTON models GAN-based, diffusion-based, and prompt-driven by embedding them into an e-commerce system. In addition, the study attempts to improve one model's quality results by applying a lightweight fine-tuning method to the challenging garment category where the model performed poorly in quantitative evaluations. By combining technical benchmarking with applied deployment testing, this dissertation aims to provide insights that are both academically rigorous and practically relevant.

The research approach follows a quantitative software-based methodology, reflecting both the technical nature of the task and the applied emphasis on system performance. This study uses the publicly available High-Resolution Zalando dataset rather than involving human participants. A controlled GPU environment which is in cloud environment of Google Colab Pro+ was used to run experiments, ensuring replicability of the research while remaining fully compliant with ethical standards, as no personal data were collected.

The remainder of the dissertation is organised into six chapters. Chapter 2 provides a literature review that critically evaluates prior research on VTON technologies, covering GAN, diffusion, and prompt-driven approaches while highlighting the persistent gap in e-commerce applicability. Chapter 3 describes the methodology, including the research design, dataset and preprocessing, model architectures, inference pipeline, and web application integration, followed by the evaluation process. Chapter 4 presents the experimental findings across both quality and performance dimensions, supported by graphical and visual examples. Chapter 5 details the web application, describing its system architecture, visualising model results, and providing deeper insights into quantitative findings. Chapter 6 analyses the experiments, comparing quantitative and qualitative results while discussing performance and scalability trade-offs. Finally, Chapter 7 concludes the dissertation by summarising the contributions, reflecting on the achievement of objectives, and proposing directions for future research and acknowledging limitations.

CHAPTER 2: LITERATURE REVIEW

2.1 Virtual try on and e-commerce

The fashion retail industry has started to show different approaches to selling their products. According to Chen et al. (2024), new digital technologies have changed the appeal of the industry. Chen et al. (2024) state that with improvements in digital technology, one-third of sales are expected to happen online. Therefore, digital technologies in this sector aim to create more interactive web pages to display products better and provide a more user-friendly experience to help consumers when deciding to buy a product. As a result, these technologies have started to emerge and improve; AI-driven recommendation systems, chatbots, and virtual try-on technologies are some examples. Kim and Forsythe (2008) agree with the idea of Chen et al. (2024) and add that the most popular and innovative one is virtual try-on models. Kim and Forsythe (2008) refer that, virtual try-on models are an interactive technology that allows users to create or use their facial characteristics, body shape, and hair colour to try the given product on their image. They also add that the product or garment used in the virtual try-on is nearly identical to the real product. In addition, Kim and Forsythe (2008) note that customer interaction and involvement can directly improve sales and the entertainment value of online shopping. These two ideas show that virtual try-on technology can enhance the online shopping experience and increase online sales in the sector.

According to Islam et al. (2024a), there are three main virtual try-on methods that dominate the literature. Islam et al. (2024c) explain that the earliest approaches to virtual try-on models were Generative Adversarial Networks (GANs), which use two neural networks to achieve a successful try-on. After that, GAN models were created and evaluated against each other by ranking the resulting image quality. This approach showed that although GAN models produced good results, they struggled with complex garments. This limitation led to the search for better models that could outperform GANs, which gave rise to diffusion models. Diffusion models, which are newer and more effective than GANs, operate using Gaussian distribution sampling by iteratively denoising until successfully producing results.

More recently, the newest technologies have emerged, among them prompt-based LMM-driven virtual try-on models such as PromptDresser. According to Kim et al. (2024b), this model operates similarly to diffusion models but achieves more realistic images by using prompts and large multimodal models (LMMs). They also add that this feature can provide users with greater interactivity, as editing the prompts allows the style or other aspects of the resulting image to be rearranged. This direction resonates with the earlier arguments of Chen et al. (2024) and Kim and Forsythe (2008), who stressed that consumer involvement and interactive experiences are central to enhancing online shopping and e-commerce.

The following section will examine these three categories in detail and provide more information about their mechanisms, different models created under these approaches, and their strengths and limitations as reported in the literature.

2.2 GAN Based Models:

Generative adversarial networks are working with two neural networks that are trying to beat each other to make perfect results and images, in a way these two networks are like hunter and hunt. Islam et al. (2024b) refers to these two core elements of the model, which are generator and discriminator. In that case, the generator, which is the hunt in that scenario, tries to manipulate the discriminator, which is trying to learn to distinguish how the generator tries to manipulate to escape, and the discriminator will achieve which is real and fake from that experience.

This approach has limitations to control the outputs as can be seen. Therefore, according to Goel et al. (2024), conditional GAN emerged as a pivotal advancement for controlling the outputs. Goel et al. (2024) refers that, with that way more accurate synthesis will be imposed by constraints on generated images, such as class labels, poses or specific attributes like facial or clothing style. In the figure above, it can be seen how the GAN models are operating to create a virtual try-on experience.

The first approach in GAN models was VITON (Han et al., 2018). Han et al. (2018) indicated the coarse-to-fine pipeline where a clothing-agnostic person was generated, then thin-plate spline transformations warped garments onto it, and lastly the generator created the output image. According to Han et al. (2018), VITON showed the feasible approach for 2D try-on without 3D modelling. However, images were limited to 256 × 192, which is low resolution. In addition to that, it created blurry textures and logos, and it was not good for the neckline in the given garments.

The second approach was created to surpass these problems, which is CP-VTON. According to Wang et al. (2018), CP-VTON had an additional feature to solve the VITON issues. This feature was GMM, which is a geometric matching module that learned transformation parameters to improve garment alignment and better-preserve logos and textures for characteristic details of garments. With that improvement, it surpassed the metric evaluations of VITON in creating better quality results by producing sharper outputs than VITON. However, Wang et al. (2018) indicated that it was still struggling with large pose variations and occlusion.

This model-to-model improvement goes on like this, and the last and one of the newest ones is VITON-HD. Choi et al. (2021) proposed new features to make a better solution to misalignment artefacts and enabled synthesis at better resolution like 1024 × 768. According to Choi et al. (2021), VITON-HD surpassed earlier GAN models in creating better image quality, but even though there were improvements, it was still struggling with occlusions and highly detailed garments.

Model	Published	FID↓	LPIPS↓ ($\mu \pm \sigma$)
CP-VTON [34]	ECCV 2018	47.36	0.303 ± 0.043
CP-VTON+ [23]	CVPRW 2020	41.37	0.278 ± 0.047
ACGPN [35]	CVPR 2020	37.94	0.233 ± 0.047
PF-AFN [8]	CVPR 2021	27.23	0.237 ± 0.049
C-VTON	This work	**19.54**	$\mathbf{0.108 \pm 0.033}$

Figure 1: from C-VTON Source (Fele et al., 2022)

Metric	VITON	CP-VTON	VTNFP	CP-VTON+	ACGPN	CloTH-VTON
SSIM	0.783	0.745	0.803	0.816	**0.845**	0.813
IS	2.650	2.757	2.784	3.105	2.829	**3.111**

Figure 2: from Cloth-VTON Soruce(Minar and Ahn, 2020)

From the above figures it can be seen that there is a race to make the best model in the literature to create better quality images as a result. However, as can be seen from the table, other GAN models that are not mentioned above are doing the same model-to-model comparison. According to Fele et al. (2022), from the figure of C-VTON, they just evaluated their quality against other models to prove they are the best. This continues in the literature, as Minar and Ahn (2020) show the same with the Cloth-VTON figure, where they again evaluated themselves with other GAN models to prove they outperformed that model. This indicates that virtual try-on models have not been systematically evaluated or utilised within e-commerce scenarios. Moreover, there is limited evidence in the literature of efforts to improve these models with respect to their applicability in retail environments, or to critically examine their scalability, usability, and deployment feasibility. Is it scalable? Is it as real as human perception? Is it costly? Or is it well-performed? These questions were never asked to improve, rather the way followed was only how this model can be improved on the given problems. This problem will be continued in diffusion model improvement environments as well in the following sections.

2.3 Diffusion Based models:

The limitations of GANs—particularly their instability, occlusion artefacts, and difficulty preserving textures—paved the way for diffusion-based models, which generate images by iteratively denoising random noise into coherent outputs (Yang et al. 2024). Diffusion frameworks are inherently more stable than adversarial training and excel at capturing fine-grained details, making them attractive for fashion try-on tasks where garment fidelity is critical. The diffusion models occurred like this, according to Yang et al. (2024) the first proposed algorithm for the model was DDPM which is denoising diffusion probabilistic model to generate images from noise, then pseudo numerical diffusion to accelerate the generation process and latent diffusion models have been introduced to perform the process in variational autoencoder environment to synthesize better image quality. Also, Gou et al. (2023) adds that latent diffusion models are reducing the computational complexity requirements by freezing the encoder-decoder. According to Islam et al. (2024b), these diffusion models work with two Markov chains, which are forward and backward processes. Islam et al. (2024b) refers that the forward process works like manually converting any data distribution into Gaussian noise by steps, while the backward chain works like a deep neural network to decode or undo the given Gaussian noise to create an image incrementally. From the image above it can be seen how these two chains work.

This model-to-model comparison approach dominates the literature again like GAN models. Diffusion models do the same evaluation plus they evaluate themselves with GAN models as well. One of the first approaches to diffusion models was LADI-VTON (Morelli et al., 2023). According to Morelli et al. (2023), the model extended

the latent diffusion framework with two improvements. Textual inversion, which is the first one, maps the garment images into CLIP's token embedding space to achieve better textures when generating the results and the other improvement was enhanced mask-aware skip connections, which try to protect the details such as logos, seams and textures. Morelli et al. (2023) refers that compared to GAN models LADI-VTON produced better quality images, especially where GAN models are worse at preserving the details of the garment and human pose. However, there is a weakness of the model in handling extreme details such as faces, hands and complex patterned garments because the latent encoder still introduced some blur and distortion. Even though these problems exist, LADI-VTON was one of the first milestones that surpassed the GAN models.

To solve the problems of LADI-VTON and to outperform that model, StableVITON (Kim et al., 2024a) emerged. According to Kim et al. (2024a), StableVITON handled the issues of LADI-VTON, they produced zero cross-attention blocks which get clothing details and features directly into the U-Net latent layer and create better alignment. In addition to that, Kim et al. (2024a) indicate that to make better and sharper results than LADI-VTON they added attention total variation loss, which helps the model map clearly on garment regions, and they added augmentation to improve robustness. As a result, StableVITON produced sharper outputs with better-preserved logos and textures, even under more challenging backgrounds and poses. Still, although it outperformed LADI-VTON in both qualitative and quantitative metrics, it occasionally struggled with garments containing very complex prints or logo-heavy designs, where detail preservation was not always perfect.

One of the most recent diffusion models to overperform these diffusion models was IDM-VTON (Choi et al., 2024). According to Choi et al. (2024), again to solve the problem of the one behind model and outperform all other models, they produced two different sections, the first one is image-prompt adapter which makes the diffusion model follow the semantic garment descriptions and the second one is GarmentNet, which is nearly the same as what LADI-VTON used to protect details, patterns and logos. Choi et al. (2024) refers that this dual approach will quantitatively surpass other models and outperform them by creating better results. However, Choi et al. (2024) indicate that this model, using these two different sections, created more computational complexity.

Method	$FID_u \downarrow$	$KID_u \downarrow$	$FID_p \downarrow$	$KID_p \downarrow$	$SSIM_p \uparrow$	$LPIPS_p \downarrow$
VITON-HD [4]	14.64	6.10	12.81	5.52	0.848	0.1216
HR-VITON [17]	12.15	3.42	9.92	3.06	0.860	0.1038
GP-VTON [50]	10.49	2.23	7.71	2.01	0.857	0.0897
PBE [43]	15.77	6.22	14.32	5.44	0.763	0.2254
MGD [1]	13.34	3.93	11.12	3.38	0.827	0.1280
LaDI-VTON [24]	12.33	4.75	9.44	3.90	0.861	0.0968
DCI-VTON [8]	11.14	3.35	8.19	2.93	0.875	0.0816
GC-DM	9.67	**1.36**	7.11	1.12	0.862	0.0988
CAT-DM	**8.93**	1.37	**5.60**	**0.83**	**0.877**	**0.0803**

Figure 3: Cat-DM figure Source (Zeng et al., 2024)

Even though diffusion models are a powerful VTON approach to create better texture synthesis, stable and creating better adversarial artefacts than GAN models, these models are again evaluated to solve the previous model problems or to outperform the other models in the range of being the best model. Others such as

CAT-DM (Zeng et al., 2024) from the figure in that diffusion model which is not mentioned above also just evaluate and use the virtual try-on model for quality of results. However, they were never tested for e-commerce integration in terms of realism, scalability and performance which again remains unpractised and untested for e-commerce approaches.

2.4 Prompt-Driven Approach

As can be seen, GAN and diffusion models are the most used and emerged ones in the literature. However, there is a new approach that is actually using large multimodal models. According to Huang et al. (2024), large multimodal models have emerged from the rise of large language models. Huang et al. (2024) indicate that there is a need for utilizing large language models as a core part to handle multimodal tasks by stretching the single textual modality to other modalities, for example, images, audio, and video. Wu et al. (2023) add that multimodal models also help to solve the problem of traditional LLMs in understanding data other than text.

With the emergence of large multimodal models, a new opportunity was created to use them in virtual try-on. Now LMMs can understand and use images as data. Virtual try-on technology can use this to edit and create an image. With that update, one of the new models that emerged from the idea of using LMMs in virtual try-on technology was PromptDresser (Kim et al., 2024b).

According to Kim et al. (2024b), PromptDresser applies this LMM approach to tackle the unexplored use of text prompts in virtual try-on. Kim et al. (2024b) refer to a text-editable virtual try-on model that leverages LMM assistance to enable high-quality and versatile manipulation by just using text prompts. This approach, according to Kim et al. (2024b), addresses three main aspects in text-editable virtual try-on. One of them is tackling the conflict where textual information of a person's garment interferes with the generation of the new one. The second is creating data-rich descriptions on prompts for paired person-clothing data to train the model. The last one is adaptively changing the inpainting mask aligned with the text descriptions, ensuring proper editing areas while protecting the original person's form irrelevant to the new clothing.

This innovation creates a bridge between language and vision. Kim et al. (2024b) refer to this by allowing the integration of both clothing images and textual prompts into the generation pipeline. With this dual conditioning, the model captures the low-level visual details of garments and adds attributes described in the text, like long sleeve, oversized, and casual style. This shows that, according to Kim et al. (2024b), the LMM backbone of PromptDresser allows users to dynamically adjust style, fit, and garment attributes simply by editing the prompt, diffusion and GAN models cannot achieve natively. Other than that, PromptDresser overcomes some of the limitations of other virtual try-on methods, such as the interactivity problem with the user, by using style editing via prompts. While GANs and diffusion models focus mainly on garment alignment and texture fidelity, PromptDresser excels in style controllability, letting the user shape the output through descriptive text. Most importantly, it adds a new paradigm which is user-in-the-loop interactivity, while GAN and diffusion models only use images. PromptDresser also outperforms the

GAN and diffusion models in the metrics used to measure virtual try-on results.

Although virtual try-on research has made significant strides, the literature to date shows a consistent pattern which is each new generation of models is primarily evaluated in terms of image quality, without systematically addressing their scalability or applicability in commercial contexts. GAN-based methods such as VITON and its successors improved the feasibility of 2D try-on by handling garment alignment and resolution, yet they struggled with occlusion and the preservation of fine details (Han et al., 2018; Wang et al., 2018). Diffusion-based models, including LADI-VTON and StableVITON, further advanced the field by producing sharper images and more faithful semantic correspondences, while IDM-VTON introduced mechanisms to balance high-level semantics with low-level garment textures (Kim et al., 2024a; Choi et al., 2024). More recently, PromptDresser represents a shift in paradigm, incorporating large multimodal models (LMMs) to add interactivity and style control, allowing users to modify try-on results through natural language prompts. While this development suggests a new trajectory for the field, it remains focused primarily on outperforming previous models on technical benchmarks such as SSIM, LPIPS, and PSNR.

2.5 GAP in The Literature

What is largely missing from these works is a critical examination of their real-world e-commerce suitability. Online retail increasingly relies on immersive digital technologies to improve consumer confidence and reduce return rates, with some studies estimating that nearly one-third of fashion sales are expected to take place online in the coming years (Chen et al., 2024). Also, Jain (2025) adds that virtual try-on and AR technologies can improve customer engagement, helping to build loyal consumers. In addition, Nguyen et al. (2025) also add that attitude toward technology and purchase intention are positively correlated. This is justified by the fact that customers are more likely to purchase when they have a favourable attitude toward the capabilities of technology and experience its advantages while browsing online. Despite this, current VTON literature rarely evaluates factors on the e-commerce side, such as human perception, where SSIM and LPIPS can be used as perception metrics according to Zhang et al. (2018), rather than focusing only on quality. Other factors like inference time, computational cost, scalability across product catalogues, or end-user experience all of which are crucial for retail adoption are also overlooked. Instead, models are often compared only within controlled benchmark datasets, divorced from deployment considerations. This leaves a clear gap: while the technical quality of outputs has steadily improved, their feasibility for integration into online shopping platforms remains underexplored. Addressing this gap, the present paper aims to evaluate representative GAN, diffusion, and prompt-driven VTON models not only in terms of visual quality but also in relation to performance and scalability within an e-commerce system and evaluation process.

3 Methodology:

3.1 Research Approach and Design:

This paper adopts a quantitative software research approach by comparing the performance of the virtual try on models to see if these models are adaptable to e-commerce deployment. Rather than using feedback or evaluation of user, this research relies on numerical metrics which are gathered by doing experiments. This approach mainly chosen because of the nature of the research question and constraint of the study.

The research question is: Which virtual try on model offers the best trade-off between quality, performance and scalability for e-commerce integration? As a result, three different virtual try on models were implemented and compared under controlled conditions, these models were GAN-based model, diffusion-based model and prompt-based model. In addition to that, to see if these models can be improved by their weaknesses, fine-tuning approach LoRA was applied on prompt-based model to observe if there are any improvements happened in metrics which has been the worst in the group, that's why this model was fine-tuned to show if they can be improved for future references. In addition to evaluation of these models, models have been implemented in a working pipeline of a web application prototype to simulate how these models might be deployed in an e-commerce environment. This also provided additional insight into latency, scalability, usability and user experience considerations. This approach was taken because in the literature these models usually compared each other with just quality metrics, this way it can be seen that how will model can deploy, and how scalable they are, how they will perform and how will quality of results will change in that e-commerce environment approach.

3.2 Dataset and Pre-processing:

The dataset was collected from public web site Kaggle. The name of the dataset was High-Resolution VITON-Zalando Dataset. Dataset contains cloth, cloth-mask, agnostic-mask, image, image-densepose, image parse-agnostic, openpose_img and openpose_json folders. These folders contain necessary images and their pre-processed files which helped to skip making these files for virtual try on process and this dataset also chosen for three models trained from this dataset which allows the test and usage of the models easier for implementation and integration. After that, dataset was augmented into four different categories, which are normal garments, logo-bearing garments, textured garments and extreme garments which are just tops. The reason why these four garment categories were selected is because they represent the most challenging and important visual features consumers consider during online clothing purchases clarity of colour, brand/logo visibility, texture fidelity, and fit versatility. The whole dataset was uploaded to Google Drive, due to size of the dataset and outputs and Colab environment can easily access the Google Drive. This improves the persistency across the sessions, allowing all models to access the dataset easily from the cloud environment of Google Drive. Also, reason why this dataset has been chosen is because the models that are used in this project are trained with it

which will give better fairness for evaluation when it is compared with same dataset and dataset is also generally used one in literature to compare models. However, one limitation of this dataset is there are limited garments and human poses which will create a little bit bias because it is not that huge for evaluating whole world's fashion environment in e-commerce.

3.3 VITON-HD (GAN-based)

VITON-HD was selected as the representative GAN-based paradigm due to its high-resolution synthesis capabilities and widespread adoption in virtual try-on research. The model builds upon a conditional GAN framework, where inputs include a person image, garment image, and supporting structural cues such as pose maps and human segmentation. A key innovation of VITON-HD is the use of ALIAS Normalization, which mitigates spatial misalignment between the warped garment and the target body representation, allowing for more realistic integration of clothing items. The pipeline operates in two stages: first, generating a coarse clothed-person representation guided by pose and segmentation; and second, refining this output through GAN-based synthesis to achieve high-fidelity textures and sharper details. Unlike earlier VTON systems limited to low resolutions, VITON-HD produces images at 1024×768 resolution, making it particularly suited for scenarios that demand both visual quality and efficiency in e-commerce applications (Choi et al., 2021).

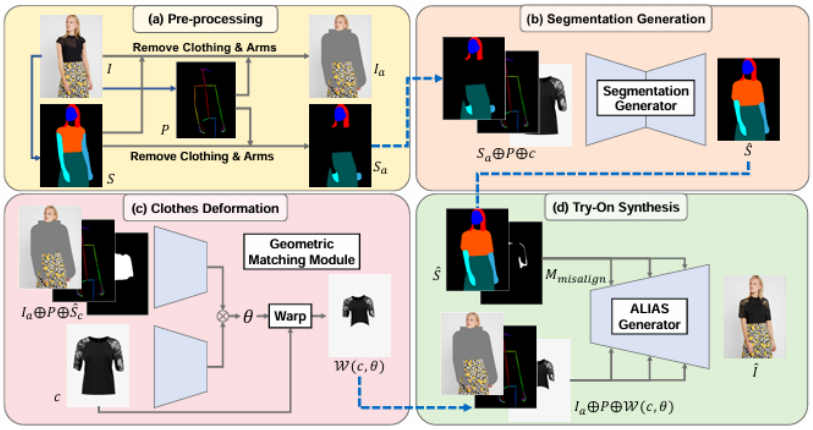

Figure 4: VITON-HD operational image Source (Choi et al., 2021)

3.4 IDM-VTON (Diffusion-based)

IDM-VTON was selected as the representative diffusion-based paradigm for this study. It extends latent diffusion models to generate high-fidelity try-on images by combining multiple conditioning signals. The framework integrates three primary components: TryonNet, a UNet backbone that processes latent person representations guided by segmentation and DensePose maps; the Image Prompt Adapter, which uses a frozen CLIP encoder to embed garment semantics into the diffusion process; and GarmentNet, a parallel UNet encoder that preserves low-level garment details such as logos and textures. In addition, IDM-VTON employs detailed garment captioning (e.g., sleeve type, neckline, style) to improve semantic conditioning, and a lightweight customisation module that fine-tunes decoder attention layers for unseen garments.

This hybrid design allows IDM-VTON to achieve superior fidelity and authenticity compared to prior GAN- or diffusion-based VTON methods, particularly in preserving complex garment patterns (Choi et al., 2024).

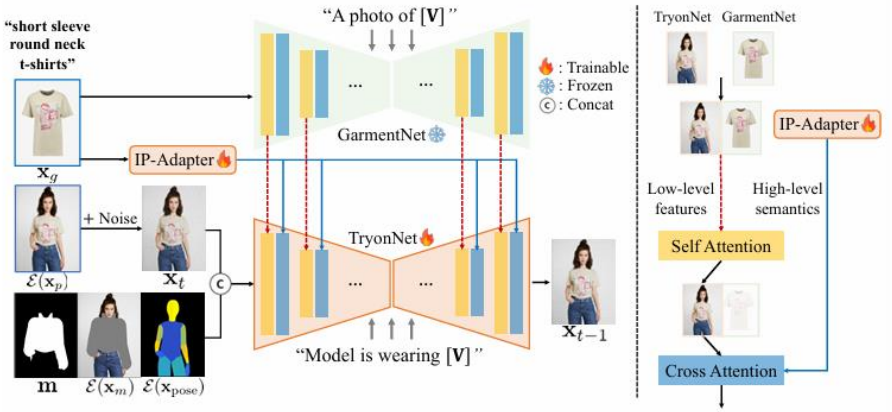

Figure 5: IDM-VTON operational image Source (Choi et al., 2024)

3.5 PromptDresser (Prompt-driven)

PromptDresser represents the prompt-based paradigm in this study, introducing generative textual prompts into the virtual try-on process. Unlike traditional VTON models that only rely on person and garment images, PromptDresser leverages large multimodal models (LMMs) to generate rich attribute descriptions of both the clothing and the person, such as sleeve length, neckline, fit, and pose. These descriptions are encoded as prompts and injected into the generative pipeline, enabling fine-grained control over the output (Kim et al., 2024b).

Architecturally, PromptDresser builds upon a latent diffusion backbone and incorporates a dual U-Net framework: a frozen reference U-Net to preserve garment detail, and a main U-Net that reconstructs the try-on image conditioned on textual and visual inputs. To address alignment issues, the model employs a prompt-aware mask generation (PMG) strategy with random mask augmentation, improving garment placement and flexibility compared to clothing-agnostic masks.

Through this design, PromptDresser achieves competitive visual fidelity while uniquely enabling style-conditioned outputs. This controllability offers clear potential for e-commerce, where consumers benefit from viewing garments under multiple stylistic perspectives in a single interface.

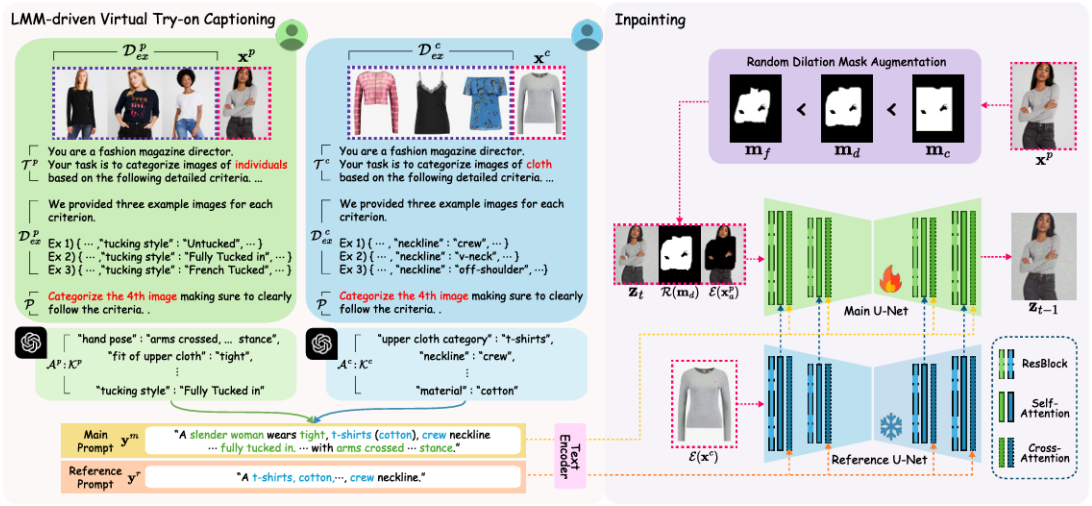

Figure 6: Prompt Dresser operational image Source (Kim et al., 2024b)

3.6 LoRA Fine-tuned model:

LoRA fine-tuning was applied to the PromptDresser model because it consistently produced weaker results on textured garments, a challenging category for e-commerce scenarios where fabric details strongly influence consumer trust. The aim was to demonstrate that such models can be incrementally improved for practical retail deployment. LoRA was chosen specifically because it is parameter-efficient, allowing effective adaptation using limited data and computational resources (Shuttleworth et al., 2024). Given the large size of PromptDresser, this lightweight approach enabled targeted fine-tuning without prohibitive training costs, aligning with the study's focus on scalable, real-world solutions.

Implementation of LoRA fine-tuning was done by three different files, where it used training, YAML and dataset files to train the model successfully. Instead of adapting all attention layers, only mid and upper block k, q and v projections were partially tuned, which helped balance memory usage efficiency and stability while leaving some layers unchanged. This parameter-efficient choice improved detail preservation at low cost but could not fully resolve the model's deeper architectural limitations. Even this approach created some VRAM problems, that's why in this training it used just upper and middle layers. LoRA was selected because it is more efficient and scalable than normal fine-tuning and also smaller, but it is not that effective compared to normal fine-tuning.

The reason these models were selected is that VITON-HD, IDM-VTON, PromptDresser, and its LoRA fine-tuned variant represent recent state-of-the-art approaches within their respective paradigms (GAN, diffusion, and prompt-driven). In addition, all of them were originally trained on the chosen Zalando dataset, which ensured a fair and consistent basis for evaluation in this project's e-commerce scenario. Nevertheless, practical implementation revealed certain challenges: the codebases and required libraries had evolved since their original publications, creating discrepancies between reported results and reproducibility in a Colab environment. Furthermore, while these models are representative exemplars, they cannot capture the full spectrum of existing approaches in the literature. Additional comparative

studies across a wider range of models would therefore be necessary to establish stronger and more generalisable conclusions.

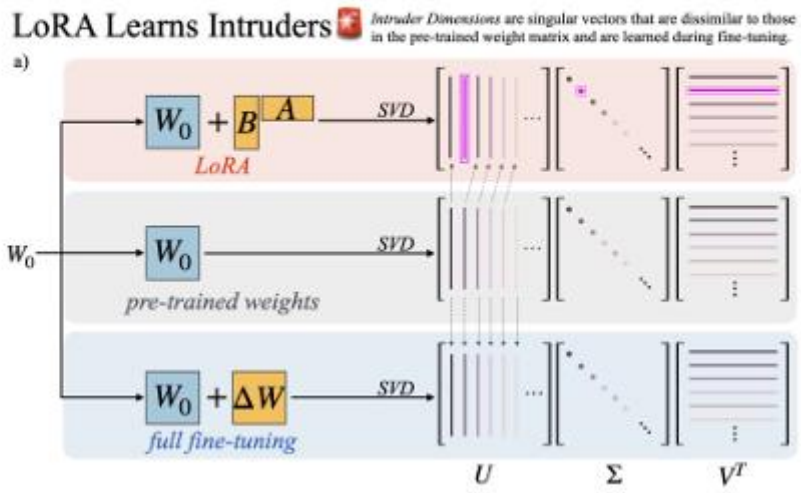

Figure L: LoRA fine-tuning vs Fine-tuning Source (Shuttleworth et al., 2024).

3.7 Inference Pipeline and Web Application Integration:

All models, fine-tuning, backend, inference and evaluation codes were conducted using Google Colab Pro+. The reason that Colab Pro+ was used is because the models are heavy, and they need a better hardware environment, so with that A100 GPU, T4 GPUs with high VRAM environment can be accessed in Colab Pro+. The local machine of the research has been used for front-end development. The libraries that have been used in this Python environment were NumPy, PyTorch, HuggingFace, Diffusers … and many more to successfully operate the models one by one in this environment.

For the backend code FastAPI has been used for creating APIs for calling the models to start inference of the models. These FastAPI endpoints have different methods for all models because they have different needs while running their inference codes. There was an issue with connecting the backend with frontend because the backend code was on the cloud and the front-end code was on the local side. To solve that issue ngrok was used to create tunnelling between frontend and backend. These approaches are very effective to make the backend and frontend connection fast and efficient; however, in deployment environments other approaches can be used like Docker or cloud which make the backend code ready and usable for all environments.

For the front-end code Vite-React has been used to create a user-friendly web application which is nearly a simulation of one of the biggest e-commerce websites, Amazon. Amazon also has a user-friendly UI/UX, and for the clothing part it has different angles of the images where clothes are worn by the model and shown in different styles and angles. This idea has been used for three different virtual try-on

models to create their try-on images with the three different try-on results. Also, the PromptDresser style feature has been used for adding more perspective to the user who is going to use this website, where there are three different images for three different styles created and shown to the user as well.

The web application was developed as the final stage of the pipeline, enabling users to upload their images, select a preferred VTON model, and visualise the corresponding try-on outputs. In addition, consumers could explore three stylistic variations of the same garment (e.g., casual, business, sporty) before making a purchase decision. The application thus functioned as a proof of concept, demonstrating how advanced VTON models can be embedded into e-commerce workflows. Importantly, it allowed the evaluation of models not only through quantitative image quality metrics but also in terms of deployment feasibility, latency, scalability, and user experience. By embedding the models within an interactive interface, the study simulated a realistic online retail environment, thereby bridging the gap between theoretical performance evaluation and practical applicability.

3.8 Evaluation Process:

The evaluation process was conducted entirely within the Google Colab Pro+ environment, as the backend code and model checkpoints were hosted there and GPU resources were available for inference. To ensure fairness, all three models were evaluated under identical conditions using the same person–garment pairs, with inputs standardised to 768×1024 resolution. A total of 400 test cases were prepared, divided equally across four garment categories: plain garments, logo-bearing garments, textured garments, and extreme garments (e.g., hoodies, tank tops). These categories were selected because they represent common yet challenging cases in online retail, where consumers place particular importance on brand visibility, texture fidelity, and versatility of garment fit.

For each model, the evaluation pipeline was automated via a custom Python script which executed inference across the test set and calculated both image quality metrics and system performance metrics. These metrics were used because quality metrics that evaluated the models were used in the literature, and in this project, they were used again for comparing the results created from the models to see how successful these metrics calculate the quality and human perception, as well as where human perception is used as human evaluation. Other than that, performance and scalability metrics were chosen for real-world e-commerce scenarios to show how user and company or retailer experience evaluation, which were all integrated into the web application to visualise better. Test cases were pre-written as text files mapping person and garment images, ensuring consistent reproducibility across models. This setup enabled a controlled comparison of outputs and provided a reliable foundation for subsequent analysis.

4 Experiments and Results:

4.1 Introduction and Evaluation of Metrics:

This chapter will present the results of the tests of four different virtual try-on models. Evaluation has been done in four different categories of garments: plain normal clothes, logo-heavy garments, textured garments, and extreme cases such as hoodies and tank tops that are different from normal tops. These four different categories are tested with two different approaches. The first one is quality metrics, which are SSIM, PSNR and LPIPS. These metrics have been selected mainly because there is no human evaluation in this project, and these metrics can evaluate human perception according to Nilsson and Akenine-Möller (2020), Keleş et al. (2021), and Zhang et al. (2018). This will help us to evaluate a scenario where humans are using this web application, and when they see the results of these three different models, how they will react will be calculated by these metrics, which will help the e-commerce alignment in that case. The other approach was performance of the models, which are inference time, model size, GPU memory usage, and output file size. These were selected because they will give a hint about, when the web application is deployed, how scalable the models are, how fast they can generate images, and how much it will cost for the company side. As a result, these two different approaches will give insights about both users and companies which will use this e-commerce website to sell their garments.

4.2 Quality Metrics:

The Structural Similarity Index (SSIM) is a perceptual metric designed to evaluate image quality by comparing luminance, contrast, and structural information between two images and has been shown to have a stronger correlation with human visual perception than traditional error-based metrics (Nilsson & Akenine-Möller, 2020). If it is close to 1 it is better.

Peak Signal-to-Noise Ratio (PSNR) is a fidelity metric that comes from MSE (mean squared error), which is expressed in decibels and normalises the reconstruction error against the image's dynamic scale. This allows the evaluation of restoration and compression of the image quality (Keleş et al., 2021). If it is the highest, it is better.

The Learned Perceptual Image Patch Similarity (LPIPS) metric compares deep features that are extracted from pretrained neural networks to find out perceptual similarity between images, which is closely related to human judgements of visual quality (Zhang et al., 2018). If it is close to 0 then it is better.

4.3 Performance Metrics:

Inference time refers to the latency experienced by the system to generate an output image, measuring responsiveness a critical performance metric in visual applications. GPU memory usage captures the peak VRAM required during inference, which signals hardware feasibility and deployment cost. These performance metrics

have been shown to correlate strongly with practical usability in real-world applications, where inference delay and hardware demands can often dominate system design decisions (Wang et al., 2025).

Model	Category	SSIM	PSNR	LPIPS	Resolution	InferenceTime	MemoryUsage	ModelSize	OutputFileSize
VITON-HD	Normal	0.8119000	14.47000	0.2220000	768 x 1024	0.877400	623.300	1176.500	0.04700000
IDM-VTON	Normal	0.8045000	14.55000	0.2180000	768 x 1024	6.490000	459.400	54344.200	0.04500000
PromptDresser	Normal	0.7986035	14.19015	0.2224642	768 x 1024	5.726790	3397.135	9794.378	0.04572680
VITON-HD	Logo	0.8118000	14.12000	0.2315000	768 x 1024	0.906100	622.900	1176.500	0.05000000
IDM-VTON	Logo	0.8103000	14.20000	0.2267000	768 x 1024	6.554500	458.900	54344.200	0.04800000
PromptDresser	Logo	0.8002054	13.92922	0.2224672	768 x 1024	5.733631	3134.469	9794.378	0.04919022
VITON-HD	Texture	0.7837000	15.04000	0.2424000	768 x 1024	0.335300	622.800	1176.500	0.06600000
IDM-VTON	Texture	0.7763000	15.05000	0.2441000	768 x 1024	6.520700	459.300	54344.200	0.06500000
PromptDresser	Texture	0.7623536	14.32196	0.2413154	768 x 1024	5.637700	3437.675	9794.378	0.06971742
PromptDresser	Texture	0.7624812	14.29294	0.2402751	768 x 1024	6.005393	4693.498	9794.378	0.06978072
VITON-HD	Extreme	0.8116000	14.88000	0.2175000	768 x 1024	2.846200	633.100	1176.500	0.05200000
IDM-VTON	Extreme	0.8105000	15.06000	0.2132000	768 x 1024	6.577500	459.000	54344.200	0.05000000
PromptDresser	Extreme	0.7991057	14.64465	0.2195753	768 x 1024	5.603473	3816.214	9794.378	0.05167728

Figure 7: **Overall Results with Lora Results (Red labelled).**

Figure 8: Bar Graph of SSIM

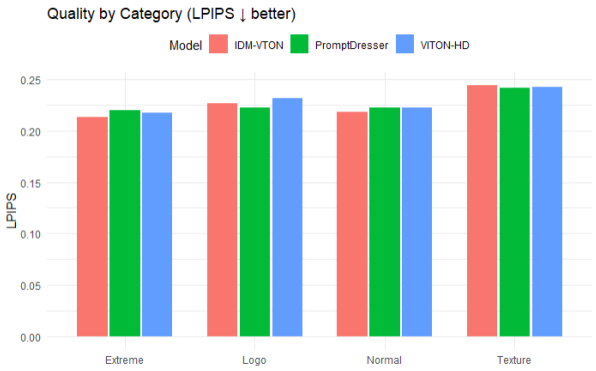

Figure 9: Bar graph of LPIPS

Figure 10: Bar Graph of PSNR

Figure 11: Normal Garment Samples

4.4 Plain-Normal Garments:

From Figure 8 it can be seen that in normal garment tests VITON-HD achieved the highest SSIM result and from Figure 7 the actual real value is 0.81, which may show good structural preservation by being close to 1. Other than that, according to Figure 9 and Figure 10 PSNR and LPIPS, IDM-VTON achieved better results with 14.55 on PSNR and 0.2180 LPIPS, which needs to be close to 0 for the best result. Lastly, PromptDresser got the worst quality results according to Figure 7, where it only achieved 0.79 SSIM, 14.19 PSNR, and 0.22224 LPIPS.

Despite strong metric performance of VITON-HD, qualitative inspection tells a different story, as shown in Figure 11. The VITON-HD result appeared blurred with oversmoothed garment edges. On the other hand, PromptDresser created better results than VITON-HD visually, but the metrics tell a different story as said before. This might show that metrics overstated perceptual realism, particularly for GAN-based synthesis.

Figure 12: Logo-heavy Garment Samples

4.5 Logo-Heavy Garments:

From Figure 8 it can be seen that VITON-HD got the better SSIM with 0.811 in logo-heavy garments, IDM-VTON followed it with 0.810 and PromptDresser with 0.80. For PSNR, from Figure 10 and Figure 7 it can be seen that IDM-VTON got the better result with 14.2, followed by VITON-HD with 14.12 and lastly PromptDresser with 13.92. For LPIPS represented in Figure 9 it can be seen that PromptDresser got the better result by being closer to 0, with an actual value of 0.222, then IDM-VTON with 0.226 and lastly VITON-HD with 0.23.

However, again VITON-HD has amazing quality metric results, but the result images show differently. From Figure 12 it can be seen that PromptDresser successfully preserved the logo of the garment better than the other two models. This

mismatch may show us limitations of SSIM and LPIPS, maybe because they were the metrics of human perception used in the project.

Figure 13: Textured Garment Samples

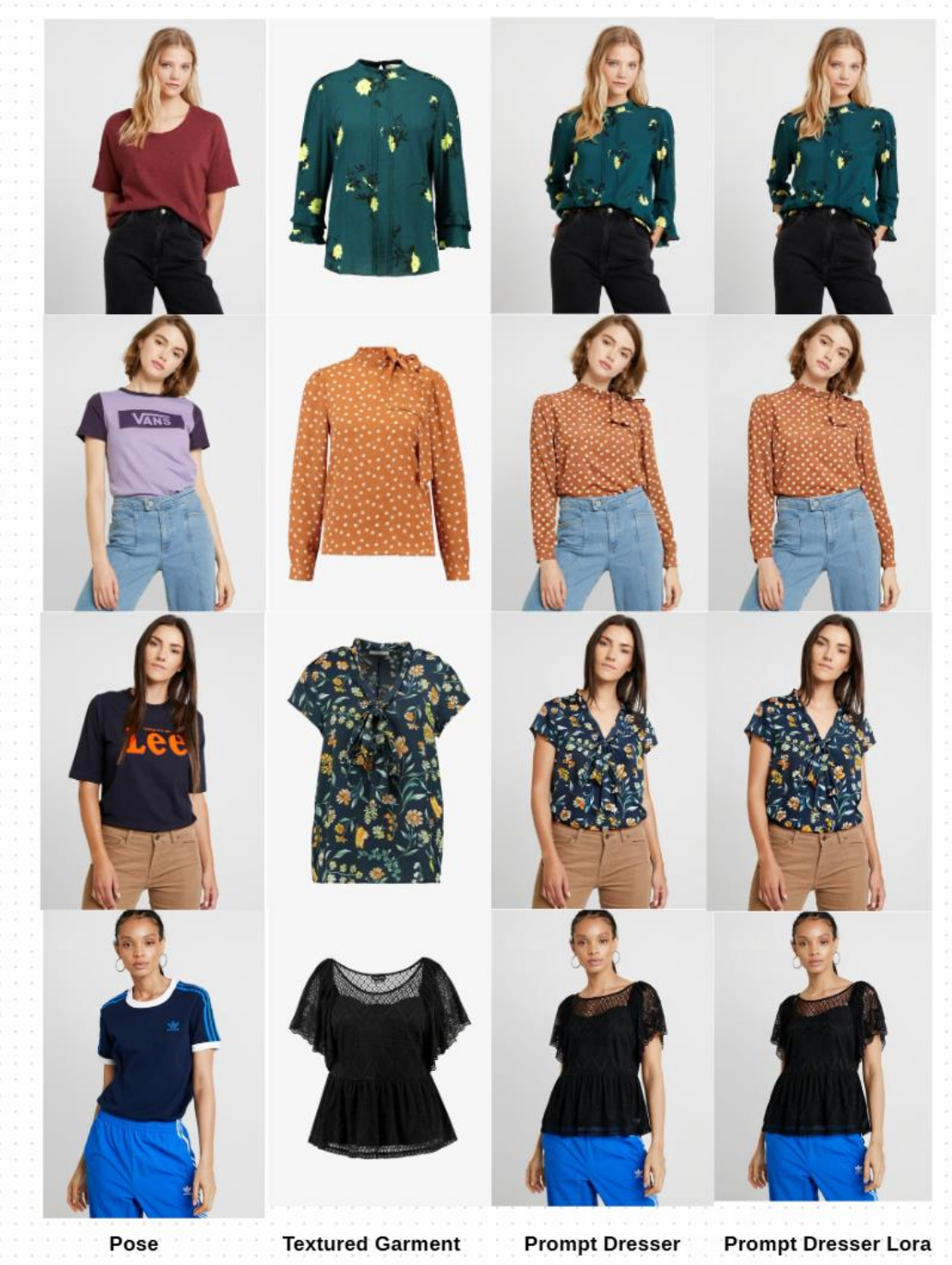

Figure 14: Lora Fine-tune Prompt Dresser Textured Garment Result Sample vs Base Prompt Dresser.

4.6 Textured Garments:

In textured garments, as can be seen in Figures 7, 8 and 9, we got the worst human perception values in the table. It may indicate that models cannot perform well

to keep up with the designs, textures, and mixed patterns of the garments. The best SSIM score was achieved by VITON-HD with 0.783, followed by IDM-VTON, then the LoRA fine-tuned PromptDresser (coloured red in the figure), and lastly the base PromptDresser.

For PSNR, from Figure 10 and Figure 7 it can be seen that the best performing model was IDM-VTON with 15.0500, followed by VITON-HD with 15.04, then the base PromptDresser and lastly the LoRA fine-tuned PromptDresser. According to Figure 7 and Figure 9, for LPIPS the best model was LoRA fine-tuned PromptDresser, then base PromptDresser, then VITON-HD and lastly IDM-VTON.

Even though quantitative results indicate that PromptDresser showed the worst performance for quality and human perception, it showed better visual results than other VTON models according to Figure 13. From Figure 7, it can be seen that LoRA fine-tuning shows small improvements. For example, the base model achieved 0.7623 SSIM and 0.2413 LPIPS, while the LoRA fine-tuned model achieved 0.7624 SSIM and 0.240 LPIPS, which are actually better but modest improved results. Also from figure 14 shows that this modest improvements in results not showing as a certain change in qualitative results.

Although, LoRA model has shown overall modest improvements both quantitatively and qualitatively, figure x shows that there is a better improvements trend on individual results, this figure x shows the improvements that better than average improvement as individual images. This reveal that LoRA achieved higher SSIM scores in approximately 46% of test cases and better LPIPS values in 41% of cases. However, Wilcoxon signed-rank tests did not reveal statistically significant differences across the metrics ($p > 0.05$), suggesting that these improvements were not consistent across the dataset.

```
        Wilcoxon signed rank test with continuity correction

data:  df$SSIM_LoRA and df$SSIM_Base
V = 2388, p-value = 0.6388
alternative hypothesis: true location shift is not equal to 0


        Wilcoxon signed rank test with continuity correction

data:  df$LPIPS_LoRA and df$LPIPS_Base
V = 2228, p-value = 0.154
alternative hypothesis: true location shift is less than 0
```

Figure a: Wilcoxon test for LoRA experiments

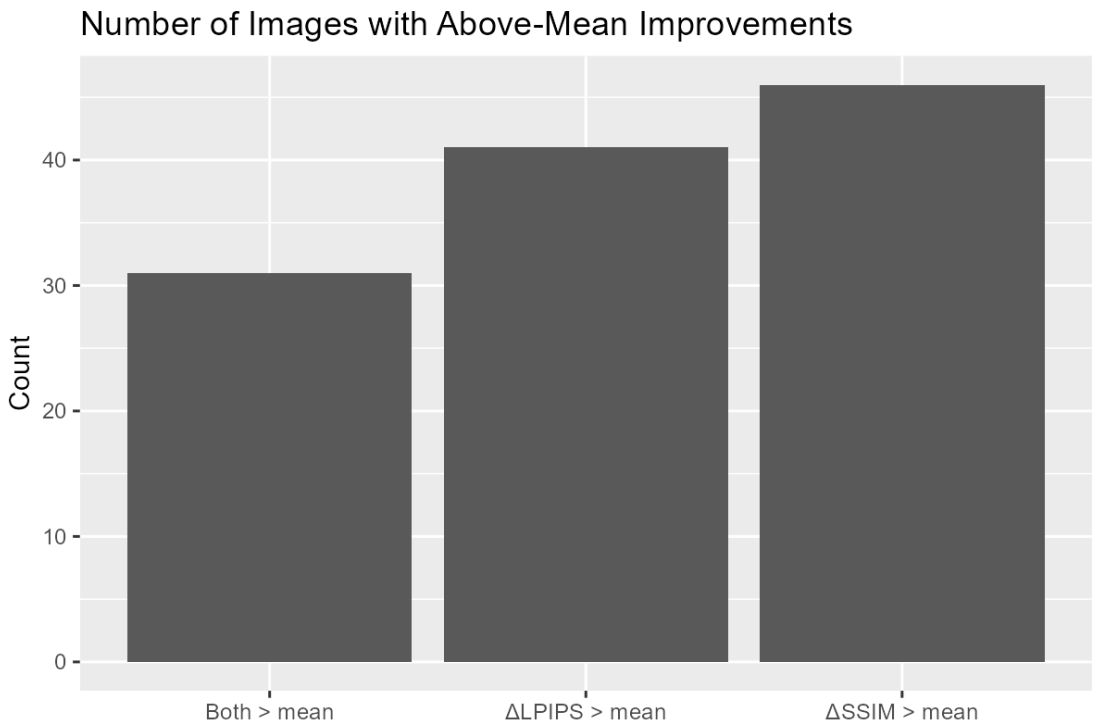

Figure x: Showing the individual improvements that are better than average improvements in LoRA

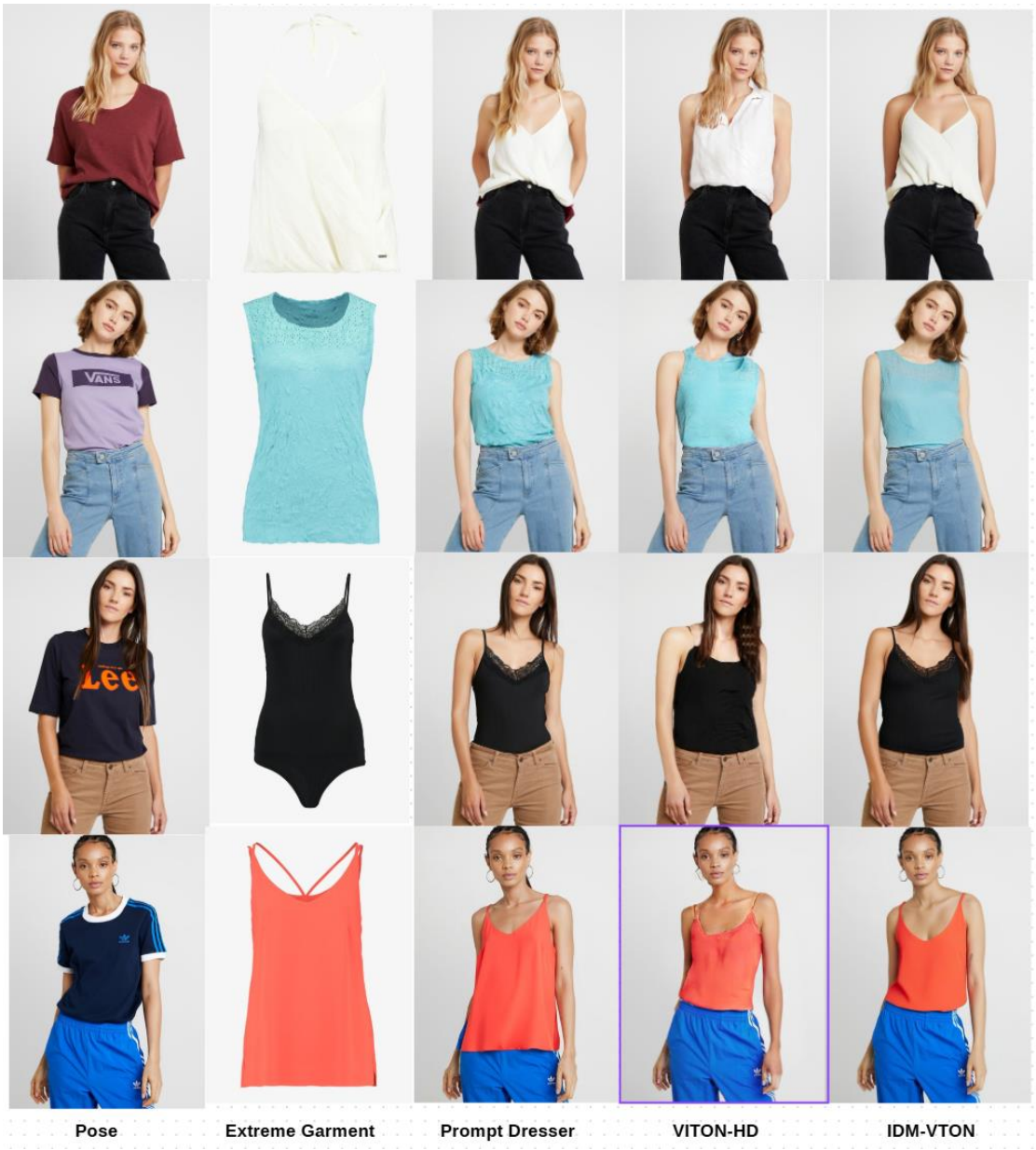

Figure 15: Extreme Garment Samples

4.7 Extreme Garments:

For extreme garments, Figures 7 and 8 show that the best SSIM value was achieved by VITON-HD with 0.811, then IDM-VTON and lastly PromptDresser. For PSNR and LPIPS, according to Figures 7, 9 and 10, the best model was IDM-VTON, then VITON-HD and lastly PromptDresser. The actual values, according to Figure 7, were: IDM-VTON 15.06 PSNR and 0.213 LPIPS; VITON-HD 14.88 PSNR and 0.2175 LPIPS; and PromptDresser 14.64 PSNR and 0.219 LPIPS.

Again, from looking at the actual visual results VTON models, in Figure 15 it can be seen that VITON-HD performed poorly compared to the other models. However, in quantitative results it achieved better results.

4.8 Performance Metrics Results:

In Figure 7 it can be seen that inference time in different sections has changed across all models. However, it can certainly be seen that the fastest model was VITON-HD, averaging 0.8 seconds per image, followed by PromptDresser with an average of 6 seconds per image, and the slowest was IDM-VTON with 6.5 seconds. This could show that diffusion models can be the slowest due to their structure, where results are processed with more steps compared to other models.

The model size shows that IDM-VTON has the largest model size at nearly 53 GB, followed by the base PromptDresser and the LoRA fine-tuned model at 9.7 GB, and the smallest model was VITON-HD, the GAN model, at 1.1 GB. This shows which models can be deployed and stored more easily for web application use.

For VRAM and memory usage, Figure 7 indicates that the LoRA fine-tuned PromptDresser used 4.4 GB, the highest of all, followed by the base PromptDresser at 3.3 GB, IDM-VTON at 0.4 GB, and VITON-HD at 0.3 GB. This shows that even LoRA fine-tuning adds additional memory usage, which could be a cost for companies wanting to improve their models.

For output size, models created larger outputs for textured garments as seen in Figure 7 (0.06 MB). This could show that models are challenged to capture textures, leading to larger outputs. Other than that, it can be seen that IDM-VTON created smaller outputs than the other models in all garment sections, making it more scalable for creating results and storing them in the database of the web application.

4.9 Summary of the Experiments and Results:

This chapter presented the results of the evaluation conducted on the selected virtual try-on models. The models were tested across four garment categories using two complementary approaches: quality metrics (SSIM, PSNR, LPIPS) and performance metrics (inference time, model size, GPU memory usage, and output size). In addition, visual outputs were inspected qualitatively to capture perceptual aspects not fully reflected in the numerical results. The models were also integrated into a web application poc to demonstrate their functionality in an e-commerce setting, next chapter will show how this web application works and why it is important to have this model. The findings from these experiments provide the basis for the analyses chapter as well, which will critically discuss and interprets the results in relation to existing literature and the overall research objectives.

5. Web Application:

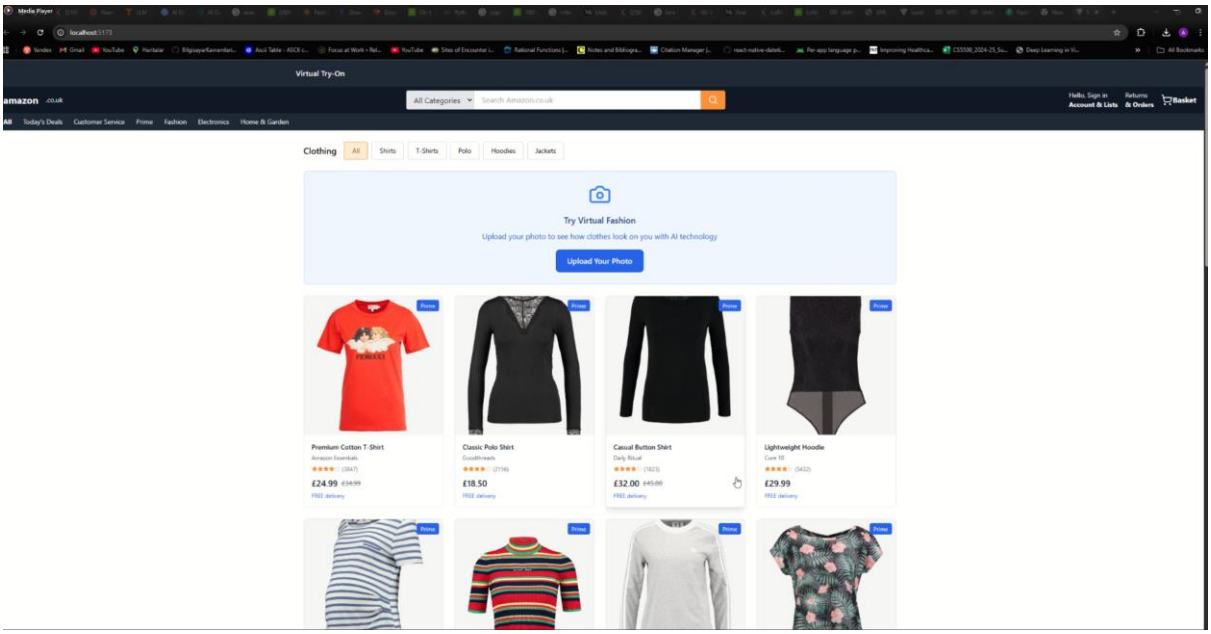

Figure 16: Web Application Main Page

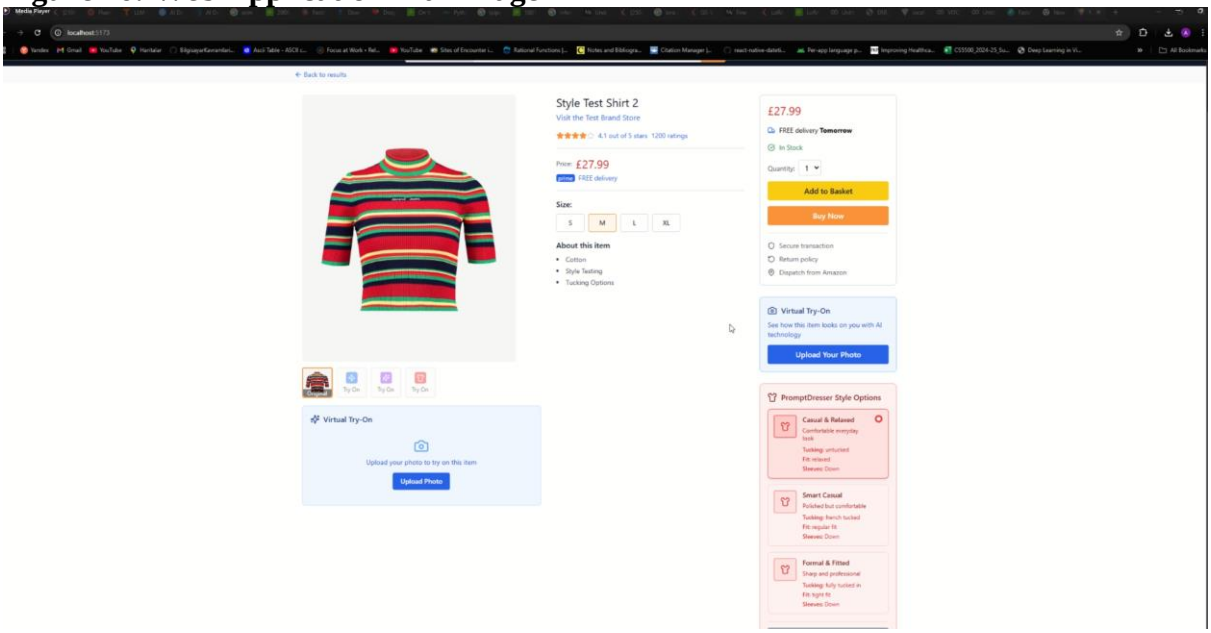

Figure 17: Web Application Detail of Product and VTON models Page

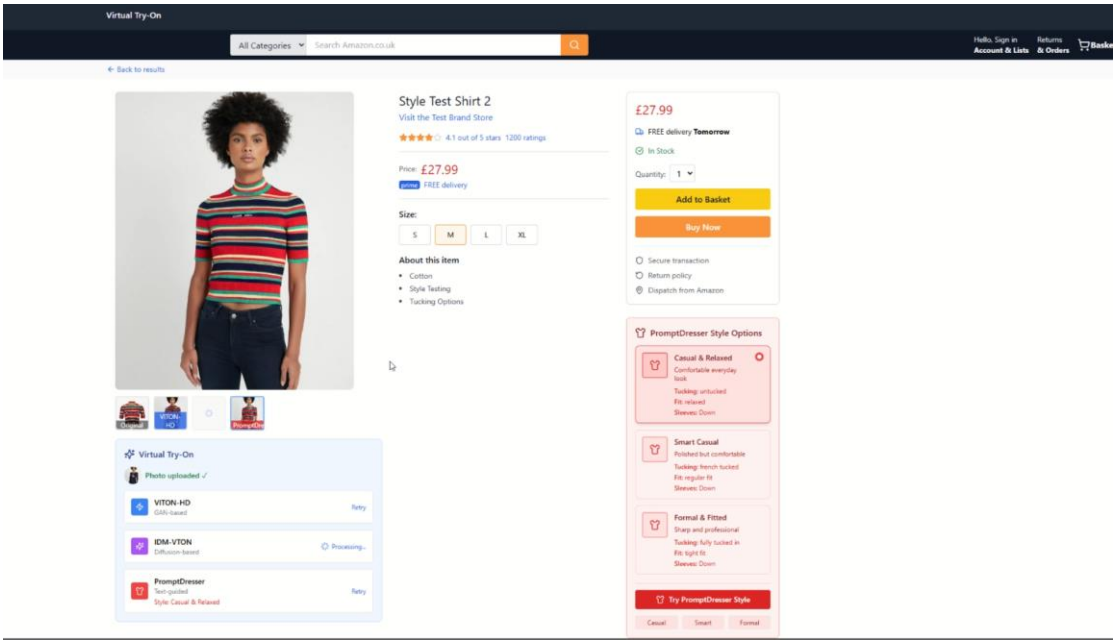

Figure 18: Prompt Dresser casual style selected where it makes the sleeves shorter and showed in web application.

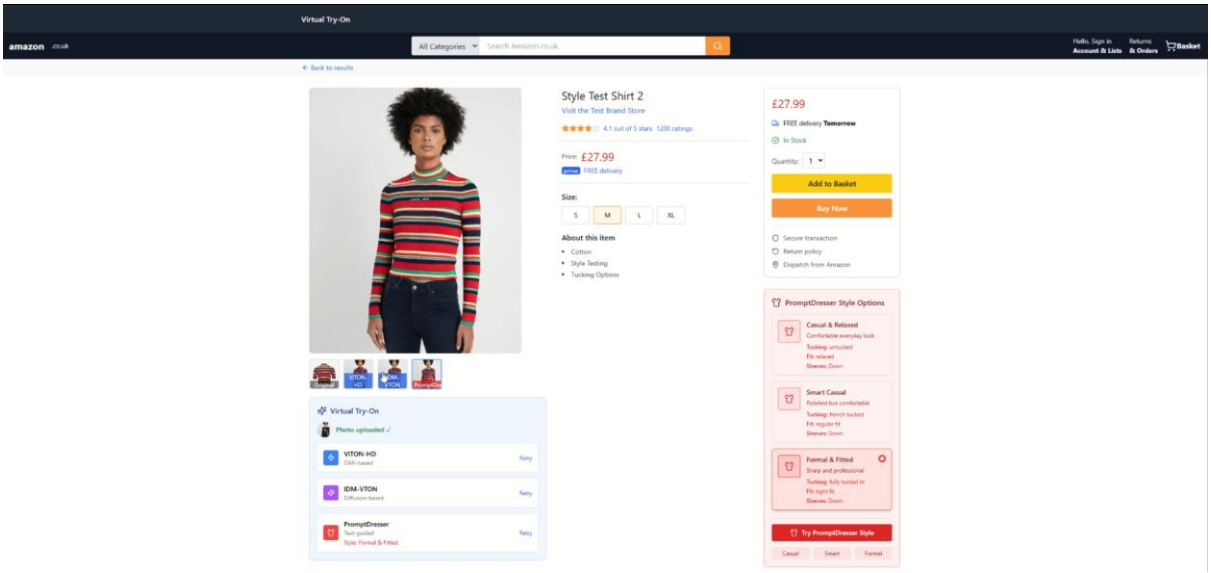

Figure 19: Prompt Dresser formal style selected where it makes the sleeves longer and showed in web application.

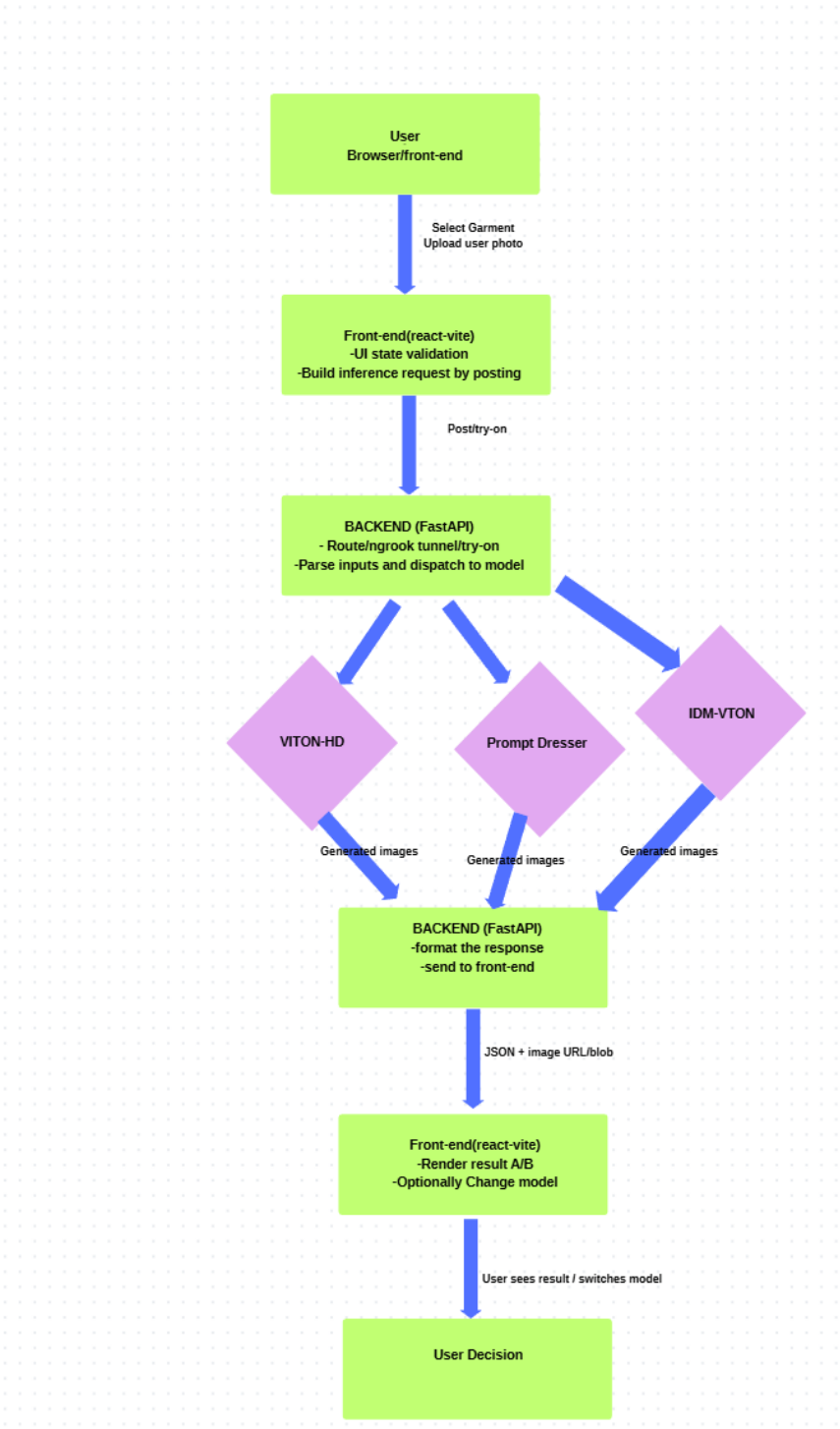

Figure 20: Web Application Workflow

5.1 System Architecture:

This web application has been created to visualise the findings it has been found, and this web application has been created to show not only evaluation has been done but practical deployment poc has been created to show as a 1 of the solutions to the gap in the virtual try-on literature.

The web application has been created around lightweight, but modular

pipeline which has divided 2 main sections, front-end and backend. The backend section was implemented in a FastAPI, which serves as a bridge for backend which consist the vton model inside with try-on API where there is inference will start when it called and this API will wait for 2 main components from the user which are 2 images, one of them is selected garment which is going to be selected on frontend and uploaded human image from the frontend again, then it will format the images as inference wants it to operate the successful Virtual try-on.

This process was doing in the colab environment which is needed because the models that it has been used in this study are very large to operate in local computer's gpu. This use of colab environment usage even it helped with operating the models successfully, there is an issue in communication between try-on API that has been created and frontend call. This issue comes from cloud environment operation, which is colab environment helping the operate the models, but this environment is cloud which cannot be called as localhost from the local computer. To solve this issue ngrook tunnel has been created in every model that has been running in colab environment which is actually nearly like deployment process but again it is private to localhost not for the other. Ngrook tunnelling is creating a link as a https like this (https://621ed3ca5a7f.ngrok-free.app), which is helping the communication between cloud environment backend API and front-end, this actually showing these model can be deployed successfully as web application not for the user only and cloud environments that are created has successfully communicate between the user interface and backend.

Front-end has been build by usage of react-Vite application, which provides a simple and typical e-commerce application mirroring. The user can browse in the garment menu , select one of them, upload their personal image or their avatar, select which model they want to use as virtual try on operation and can view their images instantly when try-on process has been finished, the can also switch or choose multiple models to try their garment on their image or avatar.

From the figure 20, it can be seen how this web application poc has been operating as a whole process. The workflow is visualising the whole process that has been doing in whole virtual try-on operation in a web application poc, it detailly gives whole operation and it separates the processing steps and the other steps by different shapes of the components in web application. Where diamond shapes as working processing steps where virtual try-on has been going on.

5.2 User Interface and Experience:

The application interface was designed like e-commerce web sites which is simulating clarity and accessibility in mind. Form the figures 16, 17, 18 and 21 it is showing the main components of the web application, which are garment or product catalogue like in e-commerce web sites, other one is showing the details of the product page, others are uploading the image, selecting the model and resulting display screen which are representing the user interactions as in real-world e-commerce web sites. Compared to simply showing raw results, this design highlights the workflow of a real-world deployment.

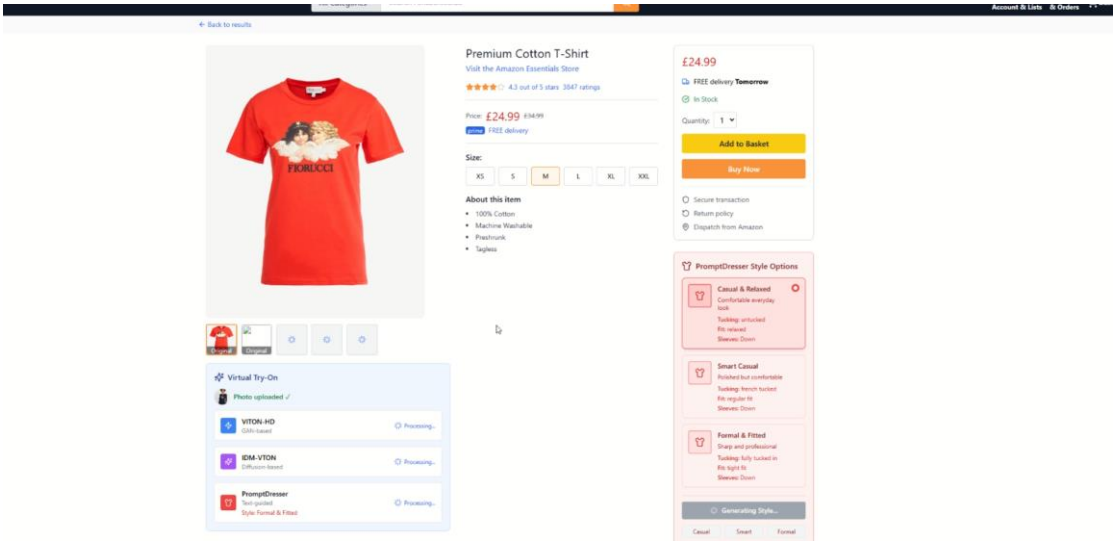

Figure 21: VITON-HD selected in the web application and result of it.

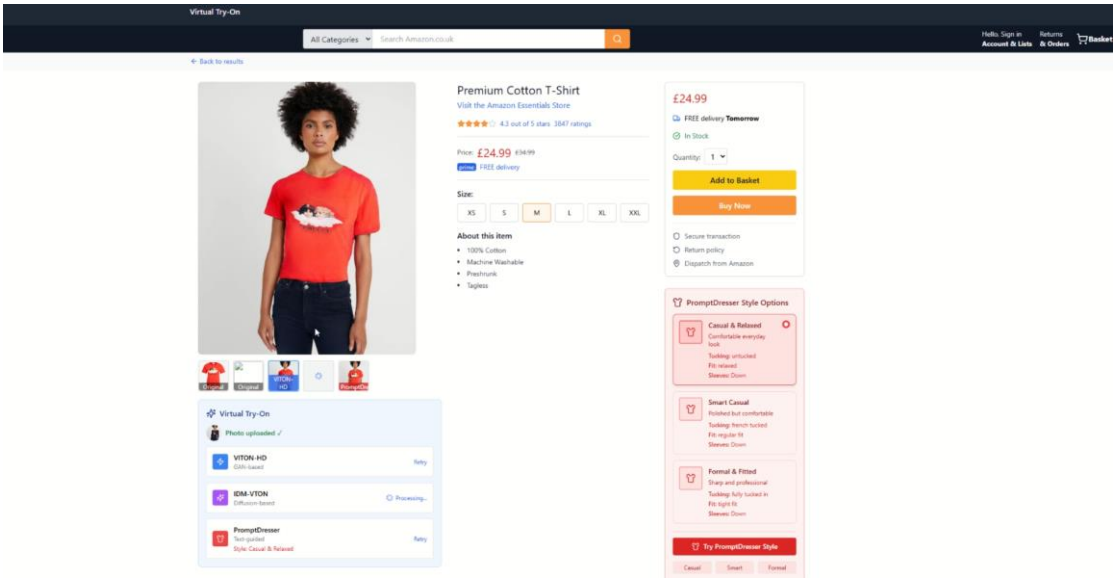

Figure 22: VITON-HD selected in the web application and result of it.

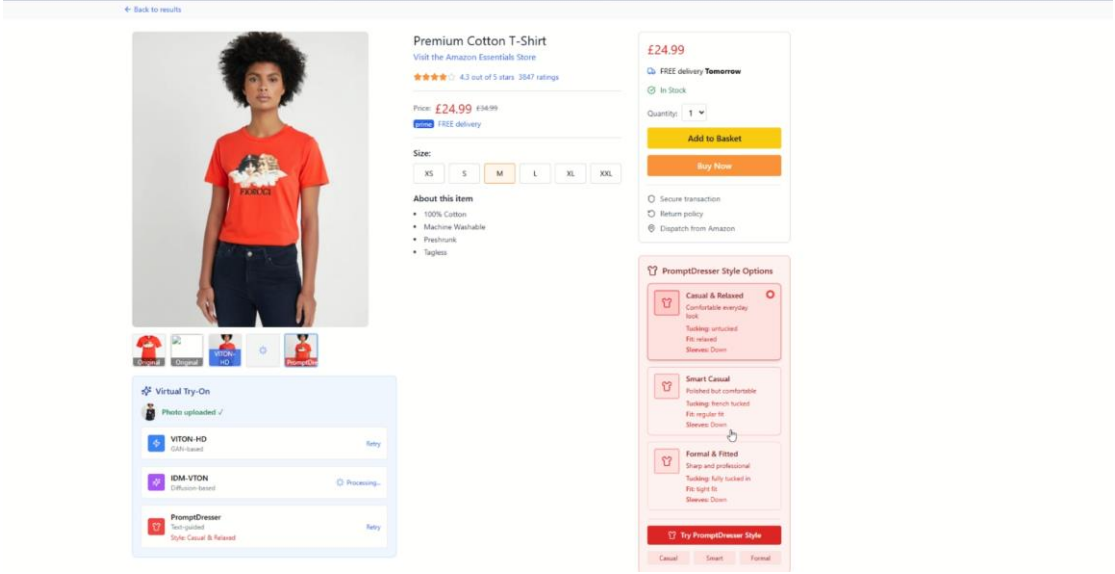

Figure 23: Prompt Dresser with casual style selected in the web application and result of it.

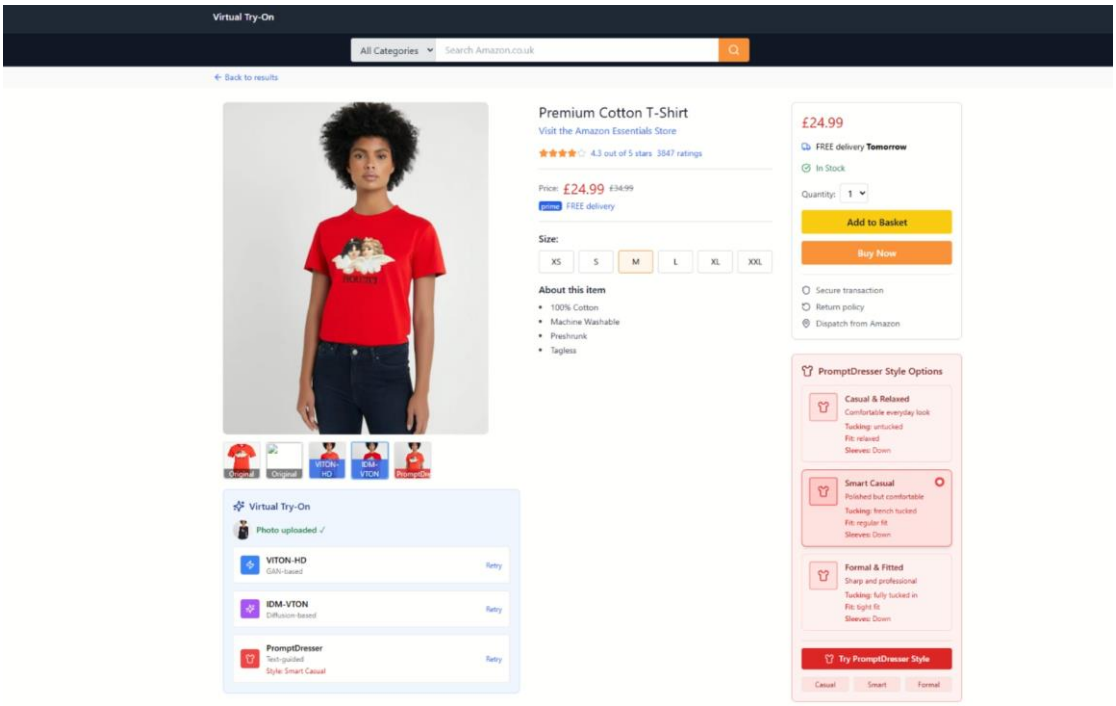

Figure 24: IDM-VTON selected in the web application and result of it.

5.3 Visual Demonstration of Results:

From the figure 22,23 and 24 also with the video (https://drive.google.com/file/d/1_8QP9qR1BlAoTiplN8Xzki0s6IMRJVlS/view?usp=sharing) of this web application while operating has showing how the models are reacting in an e-commerce scenario; form the figures it is showing the results from the models that have been created by them. The results actually align with the experiments. VITON-HD nearly instant results, however logo that is in the garment was compromised, IDM-VTON better visual quality but high waiting time and inference time for results and Prompt Dresser better visual quality, second best as inference time and from the 19 and 18 it can be seen that also provides better engagement by providing different style of selected garment.

This application reinforces the idea that numerical metrics alone cannot determine user perception. In practice, latency, detail preservation and system responsiveness all shape whether a model feels trustworthy and convenient to users. As a results, web application acted like a bridge between numerical evaluation and realistic consumer perception.

The Web application has shown that VTON models need to be operationalised beyond do model-to-model comparison just by visual quality metrics. By showing results in the operating interface, the evaluation process becomes possible not only evaluate the quality, but also latency, usability and scalability. This web application not only showing the gap of the model-to model comparison is not enough, it also work like proof of concept, showing rather than using hybrid deployment strategy to find out best model from the trade of quality, performance and scalability by

practically deployed environment.

5.4 Summary of Web Application:

In this section, one of the contributions of project has been showed, which has an impact to not only try to answer the research question but showing the practical solution to given gap in the literature by deploying and using the models in a web application as a proof of concept. The next chapter will analyse the experiments results and will show why this web application and hybrid solution that has been done in this web application is important.

6. Analyses:

Method	LPIPS ↓	SSIM ↑	FID ↓	CLIP-I ↑
		GAN-based methods		
HR-VITON [27]	0.115	0.883	9.70	0.832
GP-VTON [54]	0.105	**0.898**	6.43	0.874
		Diffusion-based metho		
LaDI-VTON [31]	0.156	0.872	8.85	0.834
DCI-VTON [9]	0.166	0.856	8.73	0.840
StableVITON [23]	0.133	0.885	6.52	0.871
IDM–VTON (ours)	**0.102**	0.870	**6.29**	**0.883**

Figure 25: IDM-VTON Source Choi et al. (2024)

Dataset	VITON-HD			
Method	SSIM ↑	LPIPS ↓	FID ↓	KID ↓
HR-VITON	**0.8767**	0.1153	12.24	3.74
GP-VTON	0.8763	0.1279	9.95	1.80
LADI-VTON	0.8630	0.1393	9.95	2.20
DCI-VTON	0.8712	0.1245	9.46	1.61
StableVITON	0.8757	0.1253	9.84	2.07
OOTDiffusion	0.8424	0.1200	9.36	1.04
IDM-VTON	0.8626	0.1023	9.20	1.27
IDM-VTON + our text	0.8650	0.1025	9.15	1.26
$\text{Ours}_{\text{pose}}$	0.8623	**0.1013**	9.15	1.21
Ours	0.8530	0.1165	**8.64**	**0.75**

Figure 26: Prompt Dresser Source Kim et al. (2024b)

6.1 Quantitative vs Qualitative Evaluation:

From the quantitative results, for the SSIM metric VITON-HD had the best SSIM score in all sections, which are normal, logo, textured and extreme garments. On the other hand, for PSNR and LPIPS scores the best model for each section was IDM-VTON. This trend and this behaviour of the models have been seen in the literature as well. According to Choi et al. (2024), from the figure 25 it can be seen that GAN models have always higher SSIM scores than other models. This behaviour also continues in the PromptDresser paper; according to Kim et al. (2024b), from Figure 26 it can be seen that again GAN models overperformed in SSIM score like the results found in this paper. Other than that, one of the other similarities of the trend caught in the results was LPIPS score; according to Kim et al. (2024b) and Choi et al. (2024), it can be seen from Figures 25 and 26 the same trend that the best score is achieved by diffusion models for LPIPS. PromptDresser is always the last model that comes in the quantitative metric comparison, where it nearly always gets the lowest score in the evaluation. Also, the LoRA fine-tuned model improved these metric results of the base PromptDresser model, but again it is not a significant improvement that can overperform the other models.

However, visual inspections of the generated results told different things. Even though VITON-HD got the highest SSIM score, it did not translate the generated image as realistically as other models. Instead, the results from the figures 12, 13 and 15 can be seen to have often blurred garment edges, loss of logos, textures and fine details. The same quality issues have been recorded in IDM-VTON images as well; even though IDM-VTON is strong in LPIPS and PSNR, from the figure 12 it can be seen that it produced logo distortions, and in another figure 13 it can be seen that textures are also a little bit distorted. On the other hand, PromptDresser produced qualitatively stronger images; from figure 12 and figure 13 it can be seen that logos and textures were preserved better than other models, even though it was the weakest according to quantitative results.

This contradiction shows that there is a metric–perception gap, where high numerical values do not necessarily equate to perceptually realistic results. This actually shows again why there is a need for real-world e-commerce suitability evaluation. As remembered, Nilsson and Akenine-Möller (2020) and Zhang et al. (2018) inferred that SSIM and LPIPS are the best metrics to evaluate human perception quality in images; however, from the qualitative and quantitative results it can be seen that there is a contradiction between them. Chilakamarthi (2011) showed that SSIM does not always correlate with human perception, particularly in cases involving blurring and smoothing. Their study highlighted that SSIM scores remain high even when human observers clearly perceive degradation. This explains why VITON-HD's oversmoothed outputs were rewarded numerically while appearing unrealistic. Also, according to Sun et al. (2023), SSIM, PSNR and LPIPS all correlate weakly with human judgments, sometimes ranking images in ways that directly contradict human evaluation. They argued that alternative perceptual metrics, such as SemSim, are needed to capture subjective realism more faithfully. Despite this, the majority of the virtual try-on literature continues to rely almost exclusively on these metrics, leaving consumer-facing realism underexplored. Our study therefore extends the field by embedding models into a working e-commerce web application, where discrepancies between quantitative scores and qualitative perception become evident in real-world contexts.

As a result, the web application that is created in this research acts like a bridge between numerical assessment and consumer experience, which is showing that virtual try-on models need to be tested at application level. Because the metrics used for quality to evaluate human perception and quality, as shown, could not be trustable when it comes to qualitative results. The web application that is created here can solve this problem if it is deployed and if the quality metrics are changed to real-world human evaluation. From the metrics that have been used to evaluate quality and human perception, it is shown that rather than choosing one model as universal, there should be a hybrid solution such as the created web application in this research; and this way, using different metrics and using different models in real-world application can show which model is the best in quality and human perception rather than the model-to-model comparison approach that is always done in the literature.

6.2 Performance and Scalability Evaluation:

While quality evaluation showed a metric–perception gap and a solution to

find the best model for quality can be usage of models as hybrid usage like the web application, performance analysis highlights a different set of trade-offs which could show which model is more scalable and can perform more than the others for e-commerce deployment. To find this out, the evaluation is going to be done on the models by inference time results, memory usage, model size and output file size.

The most major difference between models is inference time. VITON-HD, which is a GAN-based model, consistently created images in under one second. This responsiveness is very important in e-commerce platforms, where Basalla et al. (2021), infer that delays longer than 2–3 seconds increase the likelihood of user abandonment. On the other hand, IDM-VTON is much slower and the slowest model that creates the result, which is 6.5 seconds, and PromptDresser is second where there is a 0.5 second difference on average with IDM-VTON; both risk undermining user engagement, which is important according to Kim and Forsythe (2008) and Chen et al. (2024), for the usage of virtual try-on models in e-commerce as mentioned before. As a result, in a real-world e-commerce scenario a customer waiting this long for a preview image may abandon the platform entirely, diminishing the commercial utility of models with otherwise strong perceptual quality, which shows there is a trade-off between performance and quality; even if qualitatively other models create better results, GAN is much faster, which is important in e-commerce for performance of the model.

The other important aspects are model size and memory usage. Again, VITON-HD has the lightest model, which requires only 1.1 GB storage, which stands in stark contrast to IDM-VTON, which needs a 53 GB requirement. According to Barroso (2005), the importance of scaling comes from four main different components, and one of them is software infrastructure cost, which shows that if a company is going to use these models the best solution will be choosing the most scalable one, which is the GAN-based one. Other than that, PromptDresser has a 9.9 GB model size, which is the second option; the diffusion model, according to Barroso (2005), is the less scalable one to use because of the huge model size. The other main component Barroso et al (2005), refers to is the price of hardware, which comes to memory usage of models. Surprisingly, from the table it can be seen that IDM-VTON is the most scalable one that is using the least amount of GPU—an average of 0.5 GB; this can be the reason that this model is using an under thread while inferencing, that could be the reason. The second scalable model uses 0.6 GB GPU, which is VITON-HD, and the last model that is using the most GPU is PromptDresser, nearly 3.3 GB, emphasising that it is the less scalable one to use because of the most memory and GPU usage, which will cost a lot according to Barroso (2005).

Output file sizes provide an additional layer of complexity. While differences were generally small, textured garments consistently produced larger outputs, reflecting the extra computation required to capture fine detail. This reinforces the observation that computational burden is not evenly distributed but depends on garment type, adding another constraint for e-commerce platforms that must handle diverse product categories.

The role of LoRA fine-tuning shows the tension between, again, quality and scalability. Although LoRA's improvements were not statistically significant, the per-

image analysis highlights its potential value as a parameter-efficient fine-tuning method. Nearly half of the images exhibited gains in SSIM or LPIPS, suggesting that LoRA can provide occasional perceptual benefits without the need for full retraining. From a practical standpoint, this approach offers a lightweight way to adapt models to new data or garment categories, which may be especially relevant for e-commerce contexts where time and resources are limited. However, GPU consumption rose to 4.4 GB where it was 3.3 GB. This outcome highlights a recurring pattern: small improvements in visual realism can come at disproportionate computational costs. However, this improvement in quantitative results cannot be seen in qualitative results that much; even with LoRA fine-tuning, which is done with smaller hyperparameters, computational cost is increased by GPU usage, which lowers the scalability of the model in an e-commerce scenario.

Overall, evaluation demonstrates that there is a trade-off between quality and scalability and performance. This shows that in an e-commerce scenario the evaluation of scalability and performance needs to be taken with quality metrics that have been used in the literature. Because this research shows that quality metrics are not that trustable for showing the quality of the image, user evaluation is needed for the future or any new metrics like SemSim (Sun et al. 2023). On the other hand, even though qualitative results displayed that the GAN-based model has the worst results, it is the most scalable model among the others. This again shows that there should be a hybrid approach rather than choosing one model, such as the created e-commerce website—one of the solutions to show how different models can be used in an e-commerce platform.

6.3 Implications:

The findings of this study have several implications for academic research and practical software research for e-commerce. While previous research on virtual try-on is primarily focused on benchmark comparisons using just quality metrics like SSIM, LPIPS and PSNR to compare their model with others to show their results are better, these metrics are not sufficient to present the real-world viability of these systems. By revealing the gap between quantitative and qualitative perception, and by situating models within an operational e-commerce prototype, this research contributes both conceptual and practical insights that extend beyond traditional evaluation in the literature.

Academic implications are evaluating both quality, performance and scalability metrics in a model. Showing that quality metrics alone do not show which is the best model to choose; findings show that there can be differences between qualitative and quantitative results, which make the metrics that are used for human perception and quality not trustable to find the best model. Evaluating the other metrics that are related to performance and scalability is also important to find the best model. Virtual try-on models emerged for e-commerce, but these traditional evaluations are not using these models in an e-commerce platform to evaluate them; this paper evaluated the models in an e-commerce way to fill the gap in the literature and made an e-commerce web application product to show practicality of the research as software.

Practical implications for e-commerce in virtual try-on are embedding the models into an operating web application. Unlike the model-to-model comparison approach in the literature, this research evaluated scalability and performance by their cost to integrate and the mismatch between human perception and the metrics used in the literature. The web application shows the best approach to choose the best model is using them in hybrid evaluation and hybrid deployment. One business idea that can be used for this web application to deploy and use the models in an efficient way for cost–scalability–quality can be the usage of a subscription method where the GAN model can be a free tier for users to use because of the fast and less quality results, and other models can be premium subscriptions for users to use where they have more quality but less scalability. In this way, the web application can be tested, and user evaluation of these subscriptions shows what to improve in these models rather than trying to outperform models, as in the literature; this will lead to fine-tunings of the models just like in the research where the LoRA fine-tuning approach was used to improve the model to show better results to the user.

Taken together, these implications reinforce the central research question: there is no single best VTON model for e-commerce. Instead, performance and quality trade-offs demand careful alignment with deployment priorities. For academia, this requires a paradigm shift toward multidimensional evaluation; for industry, it demands flexible strategies that balance efficiency with realism. Ultimately, the study suggests that the path forward lies not in declaring winners, but in designing hybrid solutions that integrate diverse models within scalable, user-centred systems.

CHAPTER 7: CONCLUSION

This paper has investigated a research question which is, which virtual try-on model offers the most suitable trade-off between quality, performance and scalability for e-commerce platforms. By combining structural literature review, experimentation wit h models, LoRA fine-tuning and a design of functional web applications, the research provided both theoretical and practical insights for suitability of virtual try-on model technologies in retail world.

7.1 Summary of Dissertation:

Chapter 1 is introduced the problem space in virtual try on technologies by highlighting how e-commerce platforms are shaping with changes in digital world introducing the new technologies like virtual try-on, this leads to gap or problem in the literature where model-to-model comparison approach take the literature and doing just same evaluations by same metrics as image quality ones. Chapter 2 is introducing how virtual try-on model emerged and how strong they are when they used in e-commerce platforms, after that tracing these models evolution and showing the how literature use and evaluate virtual try-on models by showing strength and weakness each model by showing comparisons in themselves and try to overperform the other, thereby this motivates how this research is going to focus on that gap.

Chapter 3 is methodology part, which includes the detail of the models that is going to be used and why they selected, how the dataset selected and which dataset selected, showing what is LoRA and how it is applied why selected, the research plan in detail and the inference pipeline and integration of web application how it is done as an overview, lastly evaluation process and how it is done each methodology have been used also evaluated by their pros and cons and why used in that project, Chapter 4 is simple experiments and their result, raw results of experiments and visual representative of these results, chapter5 is detailed web application that has been created to contribute this project by helping the evaluate the experiment results and showing what solution can be done to approach the research question, how the application works with detailed workflow and images and video that is showing how this poc work.

Chapter 6 is analyses and evaluation of the results of the experiments by comparing qualitative and quantitative results and critically evaluated with literature, also in that chapter there is performance scalability evaluation as well and evaluation with literature and implication that showing how this research contributes to academically and practically, by trying to show a solution to the question evaluating the results.

7.2 Limitations:

This study has several limitations. Firstly, all tests have been done from one dataset, which is limiting the use of the whole results that are found in the context of retail e-commerce websites where they have more different garment and human images to try. Secondly, quality metrics results show that human perception results are limited to evaluate realism of the image, so user and human evaluation could be done to get better results. Third, LoRA fine-tuning is done only in one section of the model, and, because of the VRAM limitation, it is trained with smaller hyperparameters and smaller data, which prevents broader conclusions about its effect. Fourth, results were gathered on a single hardware configuration; performance may vary in different infrastructures. Finally, the web application was not stress-tested under realistic traffic conditions.

7.3 Future Recommendations:

Future research should expand the scope and address these limitations. They should test the models with more data and datasets. They should try to add more human evaluation for evaluating the quality of the image rather than using metrics like SSIM and LPIPS because of their limitations to show realism of the image. They should try to LoRA fine-tune the PromptDresser with better hardware to make training with better hyperparameters and more data, to see the better and huge difference in metrics and also in the image.

For evaluating performance and scalability they can try to deploy the web application that is created in this project, and they should try to test it with a real-traffic scenario, where they can measure the latency, throughput and consumer behaviour. Hybrid business models that split previews and premium services should be implemented and tested in real marketplaces. Longitudinal studies linking VTON adoption to return rates, purchase intent, and customer trust would provide robust evidence of economic impact.

7.4 Reflections:

From completing this dissertation, there is several strengths and weakness became apparent, strengths were ability to combine the technical experimentation with applied system development, producing both comparative insights and functional system. Another was trying new fine-tuning techniques and successfully train the model and show its results and solved the technical problems while doing that training like memory size of gpu, Weaknesses can say to be while writing and structuring this dissertation which is so stressful. Professionally this research has enhanced my confidence, because it showed me, I can do multiple work in a given little time, also boosted my knowledge about AI system designs, from model training to deployment.

REFERENCES

Başeğmez, H. & Yaman, T.T. (2022). The role of virtual try-on technology in online purchasing decision. Journal of Research in Business, 7(IMISC2021 Special Issue), pp.165–176.

Barroso, L.A. (2005). The price of performance: An economic case for chip multiprocessing. Queue, 3(7), pp.48–53.

Basalla, M., Schneider, J., Luksik, M., Jaakonmäki, R. & Vom Brocke, J. (2021). On latency of e-commerce platforms. Journal of Organizational Computing and Electronic Commerce, 31(1), pp.1–17.

Chen, C., Ni, J. & Zhang, P. (2024). Virtual try-on systems in fashion consumption: A systematic review. Applied Sciences, 14(24), 11839.

Chilakamarthi, T. (2011). Comparison between Structural Similarity Index Metric and Human Perception.

Choi, S., Park, S., Lee, M. & Choo, J. (2021). Viton-hd: High-resolution virtual try-on via misalignment-aware normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.14131–14140.

Choi, Y., Kwak, S., Lee, K., Choi, H. & Shin, J. (2024). Improving diffusion models for authentic virtual try-on in the wild. In European Conference on Computer Vision, pp.206–235. Cham: Springer Nature Switzerland.

Fele, B., Lampe, A., Peer, P. & Struc, V. (2022). C-vton: Context-driven image-based virtual try-on network. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp.3144–3153.

Goel, P., Sharma, B., Kumari, A., Gupta, A.K. & Chhajed, P. (2024). A Review on VIRTUAL TRY-ON. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp.1–5. IEEE.

Gou, J., Sun, S., Zhang, J., Si, J., Qian, C. & Zhang, L. (2023). Taming the power of diffusion models for high-quality virtual try-on with appearance flow. In Proceedings of the 31st ACM International Conference on Multimedia, pp.7599–7607.

Han, X., Wu, Z., Wu, Z., Yu, R. & Davis, L.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.

Huang, D., Yan, C., Li, Q. & Peng, X. (2024). From large language models to large multimodal models: A literature review. Applied Sciences, 14(12), 5068.

Islam, T., Miron, A., Liu, X. & Li, Y. (2024a). Image-based virtual try-on: Fidelity and simplification. Signal Processing: Image Communication, 129, 117189.

Islam, T., Miron, A., Liu, X. & Li, Y. (2024b). Deep learning in virtual try-on: A comprehensive survey. IEEE Access, 12, pp.29475–29502.

Islam, T., Miron, A., Liu, X. & Li, Y. (2024c). Dynamic Fashion Video Synthesis from Static Imagery. Future Internet, 16(8), 287.

Keleş, O., Yılmaz, M.A., Tekalp, A.M., Korkmaz, C. & Doğan, Z. (2021). On the computation of PSNR for a set of images or video. In 2021 Picture Coding Symposium (PCS), pp.1–5. IEEE.

Kim, J. & Forsythe, S. (2008). Adoption of virtual try-on technology for online apparel shopping. Journal of Interactive Marketing, 22(2), pp.45–59.

Kim, J., Jin, H., Park, S. & Choo, J. (2024b). PromptDresser: Improving the quality and controllability of virtual try-on via generative textual prompt and prompt-aware mask. arXiv preprint arXiv:2412.16978.

Kim, J., Gu, G., Park, M., Park, S. & Choo, J. (2024a). Stableviton: Learning semantic correspondence with latent diffusion model for virtual try-on. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.8176–8185.

Minar, M.R. & Ahn, H. (2020). Cloth-vton: Clothing three-dimensional reconstruction for hybrid image-based virtual try-on. In Proceedings of the Asian Conference on Computer Vision.

Morelli, D., Baldrati, A., Cartella, G., Cornia, M., Bertini, B. & Cucchiara, R. (2023). Ladi-vton: Latent diffusion textual-inversion enhanced virtual try-on. In Proceedings of the 31st ACM International Conference on Multimedia, pp.8580–8589.

Nandhakumar, C., Gayathri, P.A., Kaviya, B., Kaviya, S. & Kirupa, L. (2025). Interactive virtual fitting room: Real-time e-commerce accessories try-on with image descriptors using deep learning. In 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), pp.1148–1152. IEEE.

Nguyen, Q.H., Hanh, T.T., Hanh, N.L.M., Nhi, N.D.L., Anh, D.N. & Linh, H.T. (2025). The impacts of virtual try-on for online shopping on consumer purchase intention: The moderating role of technology experience. Cogent Business & Management, 12(1), 2500774.

Nilsson, J. & Akenine-Möller, T. (2020). Understanding SSIM. arXiv preprint arXiv:2006.13846.

None, E.J. & None, D.P.C. (2025). The impact of consumer perception towards virtual reality (try-on) features in online clothing platforms – a detailed review between e-commerce and v-commerce. Journal of Marketing & Social Research, 2(7), pp.59–64.

Shuttleworth, R., Andreas, J., Torralba, A. & Sharma, P. (2024). Lora vs full fine-tuning: An illusion of equivalence. arXiv preprint arXiv:2410.21228.

Sun, X., Gazagnadou, N., Sharma, V., Lyu, L., Li, H. & Zheng, L. (2023). Privacy assessment on reconstructed images: Are existing evaluation metrics faithful to human perception? Advances in Neural Information Processing Systems, 36, pp.10223–10237.

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.

Wang, Q., Jiang, S., Chen, Z., Cao, X., Li, Y., Li, A., … & Liu, X. (2025). Anatomizing deep learning inference in web browsers. ACM Transactions on Software Engineering and Methodology, 34(2), pp.1–43.

Wu, J., Gan, W., Chen, Z., Wan, S. & Yu, P.S. (2023). Multimodal large language models: A survey. In 2023 IEEE International Conference on Big Data (BigData), pp.2247–2256. IEEE.

Yang, X., Ding, C., Hong, Z., Huang, J., Tao, J. & Xu, X. (2024). Texture-preserving diffusion models for high-fidelity virtual try-on. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.7017–7026.

Zeng, J., Song, D., Nie, W., Tian, H., Wang, T. & Liu, A.A. (2024). Cat-dm: Controllable accelerated virtual try-on with diffusion model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.8372–8382.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E. & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.586–595.

APPENDIX A: ETHICAL APPROVAL

College of Engineering, Design and Physical Sciences Research Ethics Committee
Brunel University of London
Kingston Lane
Uxbridge
UB8 3PH
United Kingdom

www.brunel.ac.uk

2 July 2025

LETTER OF CONFIRMATION

Applicant: Mr Amil Kazimoglu

Project Title: Dissertation

Reference: 50971-NER-Jun/2025- 54357-1

Dear Mr Amil Kazimoglu

The Research Ethics Committee has considered the above application recently submitted by you.

This letter is to confirm that, according to the information provided in your BREO application, your project does not require full ethical review. You may proceed with your research as set out in your submitted BREO application, using secondary data sources only. You may not use any data sources for which you have not sought approval.

Please note that:

- **You are not permitted to conduct research involving human participants, their tissue and/or their data. If you wish to conduct such research (including surveys, questionnaires, interviews etc.), you must contact the Research Ethics Committee to seek approval prior to engaging with any participants or working with data for which you do not have approval.**
- The Research Ethics Committee reserves the right to sample and review documentation relevant to the study.
- If during the course of the study, you would like to carry out research activities that concern a human participant, their tissue and/or their data, you must submit a new BREO application and await approval before proceeding. Research activity includes the recruitment of participants, undertaking consent procedures and collection of data. Breach of this requirement constitutes research misconduct and is a disciplinary offence.

Good luck with your research!

Kind regards,

Professor Hussam Jouhara FREng

Chair of the College of Engineering, Design and Physical Sciences Research Ethics Committee

Brunel University of London

Department of Computer Science

CS5500 Ethics approval Questionnaire

Student ID: _____2449223_____

	Question	YES	NO
1	Does your project require evaluation of the web or mobile application/UI design or collection of primary data by using participants? You may meet your participants face to face, or you may conduct the research online or on the telephone. Any study that involves interaction with human participants requires ethical approval.	☐	■
2	Does your project involve you approaching a company to study their information systems or working practices?	☐	■
3	Does your project involve a high risk to either you or someone else?	☐	■
4	Could your project potentially infringe copyright laws or have patent issue?	☐	■
5	Does your project merely repeat what someone else has done?	☐	■
6	Does your project support immoral, violent, unjust or unethical activities e.g. doing a project for the military or one that may support child labour?	☐	■

APPENDIX B: Code

https://github.com/Amil1905/VIRTUALTRY-ONWEBAPP all the web application frontend, backend, evaluation and LoRA are in that repo also backend, LoRA and evaluation code is in appendix_2449223

Also, these codes like backend, evaluation and LoRA fine-tuning codes has been uploaded as a folder in a system rather than copying pasting in here.

ALL models code that is including the other sections to fire-up and environment of the models has been too big, so link provided is here showing the whole environments:

https://drive.google.com/drive/folders/1bj7fVJ5HCEm7tCgWjLsNZHo27qUC68Yk?usp=sharing -IDM_VTON

https://drive.google.com/drive/folders/1h_b97qgVyTyZH4sL8gB2xrA-kd-DaazX?usp=sharing – PromptDresser

https://drive.google.com/drive/folders/1bX1sCzftzcyNOnyF0lRDNDL43Xb5p7l0?usp=sharing – VITON-HD

Appendix C: Other

Any figures or tables that show preliminary results.