22nd International Conference on Advances in ICT for Emerging Regions (ICTer 2022): 104- 109

An Augmented Reality-based Fashion Design Interface with Artistic Contents Generated Using Deep Generative Models

Asangika Sandamini*,Chamodi Jayathilaka†, Thisara Pannala‡,Kasun Karunanayaka§,
Prabhash Kumarasinghe ¶ and Dushani Perera ‖
University of Colombo School of Computing,
Sri Lanka
Email:*2017is013@stu.ucsc.cmb.ac.lk,†2017is038@stu.ucsc.cmb.ac.lk,‡2017is062@stu.ucsc.cmb.ac.lk,§ktk@ucsc.cmb.ac.lk,
¶jpk@ucsc.cmb.ac.lk,‖udi@ucsc.cmb.ac.lk

2022 22nd International Conference on Advances in ICT for Emerging Regions (ICTer) | 979-8-3503-4613-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICTer58063.2022.10024084

***Abstract*—Fashion design is an art that reshapes the designers' imagination into visible content which requires a significant amount of time and effort. The assistance provided by the available design tools are limited in the sense of visualizing and fitting of the generated cloth on the human body. We present, ARGAN – an Augmented Reality (AR) based Fashion Design system which is able to generate a new dress when a sketch and a theme image are provided as the input into a Controllable Generative Adversarial Network. Further, this system can visualize the generated virtual 2D apparel in real-time on a real human body using Augmented Reality. To the best of our knowledge, this work is the first attempt at utilizing Deep Generative Models (e.g. GANs) in an Augmented Reality prototype in fashion designing for generate creative fashion content in 2D and exploiting the possibility of Deep Generative Models to generate fashion designs align to a theme. Our findings show that the use of the ARGAN can support fashion designers' during their designing process.**

***Index Terms*—Augmented Reality, Media Pipe, Deep Generative Models, Generative Adversarial Networks, Control GANs, Fashion Designs**

I. Introduction

Fashion designing comes with a designer's conceptual idea, or, more precisely, with designers' imagination [1]. The designer's mind is capable of generating new artistic content through his/her own imagination. Several supporting tools like Clo3D [2] and Valentina [3] have assisted the designers in drawing sketches, pattern creation, and further visualizing the dress on computer models. Apart from the designing tools, Augmented Reality also supports as a visualization medium for fashion where the real and virtual world are registered in 3D for real-time interaction [4], [5], [6], [7].

With the advancement of AI, Deep Generative Models have earned researchers' attention in generating artistic content. Deep Generative Models are studied for image-to-image translation [8], [9], text-to-image translation [10], [11] , text-to-speech translation [12], generating poems from images [13], colourizing sketches [14], [15], generating music tracks [16], editing photos [17], video generation [18] and 3D generation [19]. Additionally, Deep Generative Models are capable of providing user controllable features to the generated content [10], [15]. The main advantage of using these applications to generate creative content over human skills is that desired content can be generated in less time and less effort [20]. Since Deep Generative Models have been used in several applications based on artistic content, fashion designing can also take advantage of them.

The mentioned considerations outline that most of the tools available in generating of new artistic content are not related to the fashion domain. Furthermore, the use of Deep Generative Models for fashion design is focused on generating new content and does not provide a mechanism to visualize the generated cloth by providing a feature to virtually try on the clothes on the human body using Augmented Reality. This can significantly reduce the time taken for designers because they do not need to make the real dress in order to fit in on a model. Further, they can do several modifications to the dress in real-time and try on the model and finalize. Thus, it can be noted that an exploration of the potential in a system that provides a combination of Augmented Reality and Deep Generative Models in the fashion industry would bring more advantages to the designers.

The following are our main contributions-:

- This work provides a combined solution for synthesizing fashion from sketches and themes and visualizing on real models by assisting fashion designers in decision making.
- The research utilizes Deep Generative Models (e.g. GANs) in an Augmented Reality prototype for fashion design to generate creative fashion content in 2D.
- The research explores the possibility of using Deep Generative Models to generate fashion designs according to a given theme.

To the best of our knowledge, this is the first effort to explore Deep Generative Models for Augmented Reality based fashion designs.

II. Related Work

A. Deep Generative Models for Image Synthesis

Deep Generative Models are a class of deep models that generate new data from the same distribution when given training data. These models are capable of producing images, texts, and sounds. For image generation using latent space, GANs have outperformed VAE by generating more photo realistic, vivid images. The GAN could generate images that are closer to reality than VAE. Generative adversarial networks was first presented by Goodfellow *et al.* [21] in 2014.

1) Conditional GANs: In 2014, Mirza and Osindero proposed a variant of GAN called Conditional GAN [22] by extending the GAN framework. To perform the conditioning in the training, the condition is provided as an input to both the generator and discriminator. In 2017, Isola *et al.* [14] proposed a pix2pix, a model for image-to-image translations using a conditional GAN architecture.

2) Controllable Image Synthesis: In 2019 Lee *et al.* [23] proposed Controllable Generative Adversarial Network to control generating contents. These controlling features can be varied from classes [24], textual descriptions [11], sketches [25], colors [26] to multi-domain images [9]. Compared to conditional GAN architecture, controlGAN generates content with more details and uses an independent network named classifier to map the features into corresponding classes.

3) Texture, Color Synthesis and Style Transfer: Style transfer techniques use two images as inputs; the content image and the style [27]. The style image is the reference image and the content image is what the style should be transferred into. The idea of texture synthesis has gotten significant interest since the previous works have shown promising results from convolutional neural networks, Variational Autoencoders, and Generative Adversarial Networks. Xian *et al.* [27] in 2018, proposed a TextureGAN architecture to control the textures of objects. This study allows a user to put a texture patch into a sketch and the model generates an output with the selected texture. The researchers first trained a TextureGAN which contains a generator and a discriminator to generate synthetic images for the ground truth data set. In 2019 Albahar *et al.* [28] proposed an approach called Guided Pix2pix which is able to transfer the features such as textures and poses.

B. Virtual Try-on using Augmented Reality

Augmented Reality (AR) is defined in the Azuma *et al.* [29] survey as, "Augmented reality (AR) is a variation of virtual environments (VE), or virtual reality". In the fashion domain, technologies such as virtual mirrors [30], virtual 3D images [31], virtual fashion shows [32], virtual try-on methods [33], [34], and customized designing systems [5], [35] have been used in many studies with the rapid development of AR.

Virtual try-on can be used in the fashion industry to virtually fit clothes into a person's image. This application is considered as a graphic model and by using this application on retail shopping the customers can get a more realistic try before buying the item. Virtual Try-on consists of image-based (2D) Virtual Try-on [36], [37], 3D Virtual Try-on [38], [39], Multi-Pose Guided Virtual Try-on [40], Video Virtual Try-on [41] and Pose- Guided Human Synthesis [42]. In 2017, Han *et al.* [33] proposed an image-based virtual try-on without using any 3D information to transform the dress into another person's respective region on the body.

1) Pose Estimation: Human Body Pose Estimation is a computer vision task that is engaged with deep learning technology. Mainly the human body can model into three types as a skeleton-based model, Contour based model, and a Volume-based model [43]. In 2014, Toshecv *et al.* [44] proposed human pose estimation based on a deep neural network, and the pose estimations are calculated using DNN-based regression for the body joints. In 2019, Lugaresi *et al.* [45] proposed a framework called MediaPipe that addresses the building of prototypes and demos using machine learning algorithms to combine existing perception components. MediaPipe allows developers to build a prototype incrementally as a graph of reusable calculators. The framework can be built in any operating system that supports a GPU and it can be developed quickly and run a perception application. In 2020, Bazarevsky *et al.* [46] proposed a real-time human pose estimation detector on mobile devices using a convolutional neural network called "BlazePose". This research produces 33 body key points for a single person, and from that, they contribute to finding out a novel body pose tracking solution.

III. Methodology and Design

The high-level architecture of the proposed prototype is illustrated in Fig. 1.

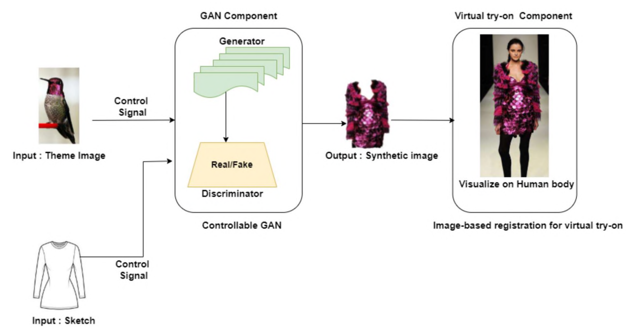


Fig. 1: High-level Architecture of the proposed work. This contains three steps - Extraction, Generation, and Visualization.

In order to generate the new cloth design using the proposed GAN architecture, we implemented a Generator (G) to create a Fashion Image (IF) from a supplied Sketch (S) and Theme (texture) Image (IT). The Fashion Image (IF) ought to be able to roughly capture the hues and textures of the Theme Image (IT). This system was trained using fashion data sets that contain 5,000 abstracts (avant-garde [47]) fashion design images. Then the resulting Fashion Image (IF) is registered onto a human body and visualized using Augmented Reality. When the model or the person in front of the camera moves, the clothing should follow accordingly to provide an AR experience. To do that a publicly available Machine Learning pipeline called MediaPipe [48] has been used and it detects the human body and its coordinates. These homography coordinates are static because the human model is moving, and the system has considered the eight coordinates of the human body as left shoulder, right shoulder, left hip, right hip, left knee, right knee, left ankle, and right ankle.

A. GAN Component

As discussed in the literature review, Xian *et al.* [27] TextureGAN architecture is the mostly related GAN architecture available for our research aim. Since their aim was also to generate synthetic natural images based on sketch and a texture.

B. Virtual Try-on Component

The extracted body coordinates for several parts of the body is mapped with the cloth image. In order to do that several steps needed to be taken.

1) Identify the video frame/surface.
2) Estimate the homography. These steps require to determine the transformation from the video frame to the target cloth image (2D). The body pose coordinates extracted from the MediaPipe steps need to considered here.
3) Apply the transformation to map the cloth image.

The homography is a 3×3 matrix. It maps the points in a plane to the corresponding point in another plane/image. The homography matrix (H) can be represented as Equation 1.

$$H = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \tag{1}$$

If the set of corresponding points of one image are (x_1, y_1) and the second image are (a_1, b_1), the Homography (H) maps them as below Equation 2

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} a_1 \\ b_1 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ 1 \end{bmatrix} \tag{2}$$

Mapping of the cloth onto the corresponding pose coordinates of the human in real-time happen as in Fig. 2.

IV. Implementation

For the implementation of the proof of concept prototype, three main components namely, generating synthetic fashion design from Deep Generative Models while controlling the extracted attributes, virtually try-on the generated 2D output registered onto the human body, and User Interface (UI) are implemented.

GAN is implemented using python. The models are trained, tested, and evaluated using a Python environment that runs entirely on the NVIDIA GPU environment and CUDA cuDNN. In the training phase, several parameters need to be fine-tuned to get the best training module.

The virtual try-on component maps the generated 2D cloth image on the human body on the camera feed. For the whole procedure, OpenCV python is used which indicates the use of computer vision-based Augmented Reality that helps in object detection.

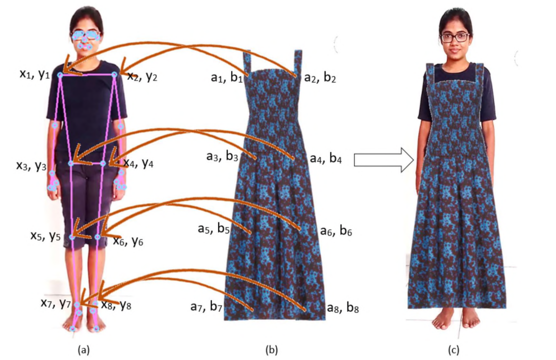


Fig. 2: The cloth pixel coordinates (a_1,b_1 to a_8,b_8) is mapped on to the corresponding pose coordinates (x_1,y_1 to x_8,b_8). (a) Human - landmarks are drawn connected to each. (b) Cloth image to be mapped on body. (c) After mapping the cloth. The arrows shows how given pixel value of the image is mapped to the corresponding region of body.

V. Evaluation

For the evaluation, the implemented GAN model was compared against baseline models using quantitative and qualitative evaluation criteria to find out which GAN model creates high-quality images.

In the user study, the subjects were presented with 5 random pair-wise comparisons and asked them to select the more realistic output with respect to the inputs and target images.

Pilot study and the main study have been conducted to measure the acceptance of the prototype by the fashion designers considering the virtual try-on components. The reason for doing the pilot study is to get ideas about the usability and user performance of the prototype. The main study was conducted after doing the relevant improvements to the suggestions and comments that were received in the pilot study.

Apart from the main study, a survey has been carried out to get feedback from the general public regarding the generated images for a given sketch and a theme input and for the virtual try-on visualization. For the survey, 5 variants of dresses that have been generated through ARGAN have been used. The participants have been asked to rate their satisfaction on a scale of highly unsatisfied, unsatisfied, average, satisfied, and highly satisfied on the design and the try-on separately for the 5 scenarios as in Fig. 3.

All of the selected participants are from the fashion design domain. And most of them are students who follow a fashion design degree and others are professional designers.

During the session, 13 questions have been asked to get qualitative answers from the participants. After the main session, open-ended questions have been asked to get feedback on the prototype.

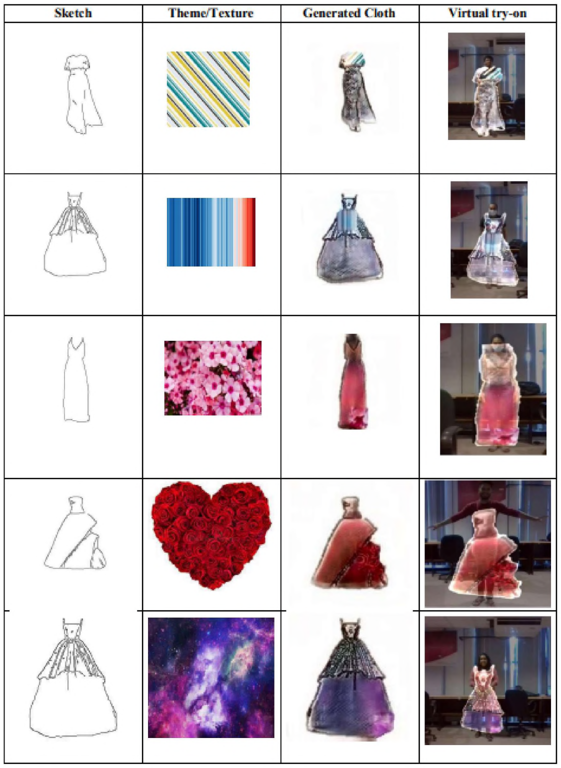


Fig. 3: Five scenarios provided for the participants in the survey.

VI. Analysis of Results

As our baseline model, we adopt Guided pix2Pix [28] which performs image-to-image translation with texture synthesis in GAN model evaluation. For comparison, both of these models are trained for our dataset. Thereafter, standard GAN performance metrics such as Frechet Inception Distance (FID) and Inception Score (IS) have been calculated using 5000 target images (real images) and 5000 output images (fake images). According to results in TABLE I lowest FID values gives the high quality of the generated samples and it proves that ARGAN method generate quality images.

TABLE I: Comparing FID score of ARGAN and Guided pix2Pix in the table FID score of the ARGAN is lower than the Guided pix2pix which represent ARGAN produce higher quality images.

Method	**FID**	**IS**	**Percentage of the User Study Results**
ARGAN (Ours)	139.180555	mean:3.74764 std-dev:0.28707448	65.6%
Guided pix2pix	239.122596	mean:4.20802 std-dev:0.17888397	34.4%

According to the survey of comparing the images in the implemented GAN model and Guided pix2pix, 65.6% users more prefer the images that generated from ARGAN method.

The results of the pilot and main study experiments are presented in both Qualitative and Quantitative manner in order to prove the hypotheses brought forward.

A. *Qualitative Analysis*

A questionnaire based qualitative analysis has been conducted after each experiment. The participants have been given open-ended questions to give their feedback on our approach. These questions have been given to measure the usability, support to the design process and the usefulness of the proposed approach. The more positive feedback have been received on the usefulness and support from the proposed approach to fashion design process.

B. *Quantitative Analysis*

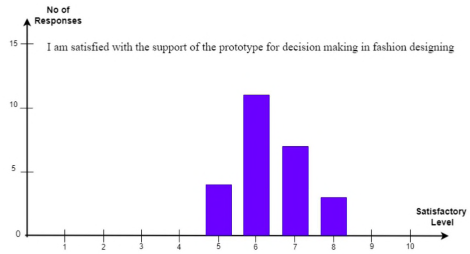
The questions which have been asked from the participants during the experiment have a satisfactory level ranging from 4-9 in the 1-10 Likert scale. From all the 13 questions, we have asked from the participants to mark their satisfactory level only for 11 questions. For those eleven questions, higher satisfactory levels have been obtained related to comfortability, usability, and the support gained from the proposed system to the fashion designers as shown in Fig. 4a., Fig. 4b., and Fig. 4c. These figures show that fashion designers are more likely to use this prototype in their design process as they can simply append their human creativity into virtual reality.

The close-ended questions were asked to retrieve direct answers from the participants regarding virtual try-on. Results obtained show that higher number of participants have been voted positively. As shown in Fig. 5. from the 25 participants, 23 of them have voted for virtual try-on using the webcam (real-time) instead of using a model video. This shows that most of the fashion designers like to try their own design in real time using Web cam rather than using a model video in the prototype.

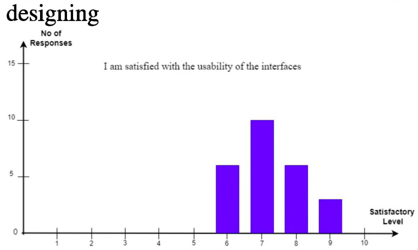
Overall, the questions asked from the participants have gained an average satisfactory level of 6.04, 7.24, 6.64 for support, usability and comfortability respectively. Therefore, it implies that people who are working in the domain of fashion design has a good impression regarding the proposed approach. Therefore it is concluded that using an interface to generate fashion design and virtual try-on would enhance the designing of cloths according to a particular theme/texture.

Furthermore, the interviews have covered all of the participants as mentioned above and 95% of the participants agreed that the prototype supports to the decision making in the design process. Most of them are satisfied with the output generated with the inputs provided by them. And 97% of the participants agreed that the little technical knowledge is required to tackle with the prototype.

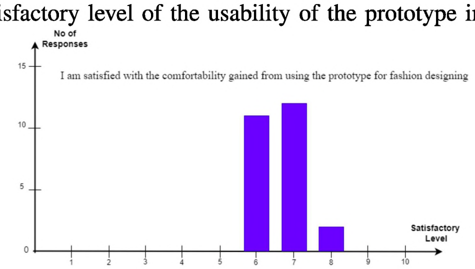
In addition to that, from the given five scenarios as in Fig. 3, the satisfactory level of the general public were



(a) Satisfactory level of the support gained from the proposed prototype for fashion designing



(b) Satisfactory level of the usability of the prototype interfaces



(c) Satisfactory level of the comfortability of using the prototype

Fig. 4: Satisfactory level

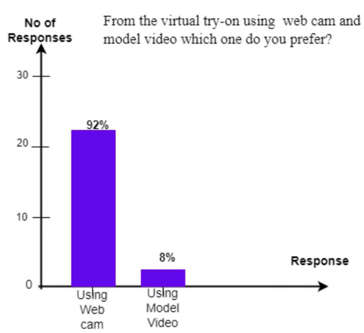


Fig. 5: Quantitative - question no.14

obtained on generated fashion image and relevant virtual try-on. From overall responses, around 75% of are satisfied with the generated output and the virtual try-on. And from other 25% responses, only around 15% are unsatisfied with the outputs and remaining 10% marked as average level.

VII. Conclusion

This research has addressed the question that "How to generate new fashion designs in 2D using Deep Generative models for an interactive designing prototype based on a given input theme image and sketch and visualize the design in real-time using Augmented Reality?".

The results obtained in the evaluation provides an idea about how well the proposed prototype supports the research objectives. The human factor and usability study conducted as the mechanism of evaluating the performance of the prototype from a designer's perspective, by measuring user satisfaction with the generated output through a questionnaire and forum. The results proves that the prototype is useful for the fashion designers in their decision making. One drawback of the usability evaluation experiment is the lack of human resources who have the knowledge in fashion domain and difficulty to engage in the study due to the pandemic. Another limitation could be identified as technical constraints related to the NVIDIA GPU environment.

In conclusion, the observations and findings from the literature review and the experimental evaluation shows that the proposed approach, can bring enhancement for the design process of fashion designers who seek an efficient way to do their designs according to conceptual ideas. Furthermore, in the Domain of Human-Computer Interaction, the proposed prototype gives a combined solution for synthesizing fashion from sketches and themes and visualizing on a real human model with Augmented Reality.

VIII. Future Work

The current study proof-of-concept prototype utilizes the generating and visualization of women's dresses only. Therefore the system can be extended for the clothing of kids and men by using relevant data sets of it. Moreover, increasing the number of data and using high-quality images for training the GAN, may result in more realistic 2D cloth images than the current prototype supports. The virtual try-on system can be improved for a wide range of movements providing more visual output to fashion designers. Not limiting to the textures or the colors of the given theme, the research can be extend toward more complex scenarios considering various aspects, shapes or behaviors of a particular theme in order to generate theme-based fashion.

References

[1] C. Eckert and M. Stacey, "Designing in the context of fashion–designing the fashion context," in *Designing in Context: Proceedings of the 5th Design Thinking Research Symposium*, 2001, pp. 113–129. 1

[2] "Clo: 3d fashion design software," accessed on 28.09.2021. [Online]. Available: https://www.clo3d.com/ 1

[3] "Download valentina program," accessed on 28.09.2021. [Online]. Available: https://smart-pattern.com.ua/en/valentina/download/ 1

[4] A. I. M. Elfeky and M. Y. H. Elbyaly, "Developing skills of fashion design by augmented reality technology in higher education," *Interactive Learning Environments*, vol. 29, no. 1, pp. 17–32, 2021. 1

[5] D. Bradley, G. Roth, and P. Bose, "Augmented reality on cloth with realistic illumination," *Machine Vision and Applications*, vol. 20, no. 2, pp. 85–92, 2009. 1, 2

[6] M. Moroz, "Tendency to use the virtual fitting room in generation y-results of qualitative study," *Foundations of Management*, vol. 11, no. 1, pp. 239–254, 2019. 1

[7] M. Kim and K. Cheeyong, "Augmented reality fashion apparel simulation using a magic mirror," *International journal of smart home*, vol. 9, no. 2, pp. 169–178, 2015. 1

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232. 1

[9] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797. 1, 2

[10] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915. 1

[11] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *International Conference on Machine Learning*. PMLR, 2016, pp. 1060–1069. 1, 2

[12] M. Bińkowski, J. Donahue, S. Dieleman, A. Clark, E. Elsen, N. Casagrande, L. C. Cobo, and K. Simonyan, "High fidelity speech synthesis with adversarial networks," *arXiv preprint arXiv:1909.11646*, 2019. 1

[13] B. Liu, J. Fu, M. P. Kato, and M. Yoshikawa, "Beyond narrative description: Generating poetry from images by multi-adversarial training," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 783–791. 1

[14] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134. 1, 2

[15] A. Deshpande, J. Lu, M.-C. Yeh, M. Jin Chong, and D. Forsyth, "Learning diverse image colorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6837–6845. 1

[16] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, "Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 1

[17] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018. 1

[18] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1526–1535. 1

[19] M. Gadelha, S. Maji, and R. Wang, "3d shape induction from 2d views of multiple objects," in *2017 International Conference on 3D Vision (3DV)*. IEEE, 2017, pp. 402–411. 1

[20] Y. R. Cui, Q. Liu, C. Y. Gao, and Z. Su, "Fashiongan: Display your fashion design using conditional generative adversarial nets," in *Computer Graphics Forum*, vol. 37, no. 7. Wiley Online Library, 2018, pp. 109–119. 1

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014. 1

[22] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014. 2

[23] M. Lee and J. Seok, "Controllable generative adversarial network," *Ieee Access*, vol. 7, pp. 28 158–28 169, 2019. 2

[24] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*. PMLR, 2017, pp. 2642–2651. 2

[25] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5400–5409. 2

[26] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-time user-guided image colorization with learned deep priors," *arXiv preprint arXiv:1705.02999*, 2017. 2

[27] W. Xian, P. Sangkloy, V. Agrawal, A. Raj, J. Lu, C. Fang, F. Yu, and J. Hays, "Texturegan: Controlling deep image synthesis with texture patches," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8456–8465. 2, 3

[28] B. AlBahar and J.-B. Huang, "Guided image-to-image translation with bi-directional feature transformation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9016–9025. 2, 4

[29] R. T. Azuma, "A survey of augmented reality," *Presence: teleoperators & virtual environments*, vol. 6, no. 4, pp. 355–385, 1997. 2

[30] P. Eisert, P. Fechteler, and J. Rurainsky, "3-d tracking of shoes for virtual mirror applications," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–6. 2

[31] M. E. Izquierdo and S. Kruse, "Image analysis for 3d modeling, rendering, and virtual view generation," *Computer vision and Image understanding*, vol. 71, no. 2, pp. 231–253, 1998. 2

[32] R. Okada, B. Stenger, T. Ike, and N. Kondoh, "Virtual fashion show using real-time markerless motion capture," in *Asian conference on computer vision*. Springer, 2006, pp. 801–810. 2

[33] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "Viton: An image-based virtual try-on network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7543–7552. 2

[34] B. Wang, H. Zheng, X. Liang, Y. Chen, L. Lin, and M. Yang, "Toward characteristic-preserving image-based virtual try-on network," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 589–604. 2

[35] Y.-P. Luh, J.-B. Wang, J.-W. Chang, S.-Y. Chang, and C.-H. Chu, "Augmented reality-based design customization of footwear for children," *Journal of Intelligent Manufacturing*, vol. 24, no. 5, pp. 905–917, 2013. 2

[36] C. Ge, Y. Song, Y. Ge, H. Yang, W. Liu, and P. Luo, "Disentangled cycle consistency for highly-realistic virtual try-on," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 928–16 937. 2

[37] K. Li, M. J. Chong, J. Zhang, and J. Liu, "Toward accurate and realistic outfits visualization with attention to details," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 546–15 555. 2

[38] F. Zhao, Z. Xie, M. Kampffmeyer, H. Dong, S. Han, T. Zheng, T. Zhang, and X. Liang, "M3d-vton: A monocular-to-3d virtual try-on network," *arXiv preprint arXiv:2108.05126*, 2021. 2

[39] I. Santesteban, N. Thuerey, M. A. Otaduy, and D. Casas, "Self-supervised collision handling via generative 3d garment models for virtual try-on," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 763–11 773. 2

[40] H. Dong, X. Liang, X. Shen, B. Wang, H. Lai, J. Zhu, Z. Hu, and J. Yin, "Towards multi-pose guided virtual try-on network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9026–9035. 2

[41] H. Dong, X. Liang, X. Shen, B. Wu, B.-C. Chen, and J. Yin, "Fw-gan: Flow-navigated warping gan for video virtual try-on," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1161–1170. 2

[42] W. Sun, J. H. Bappy, S. Yang, Y. Xu, T. Wu, and H. Zhou, "Pose guided fashion image synthesis using deep generative model," *arXiv preprint arXiv:1906.07251*, 2019. 2

[43] "A comprehensive guide to human pose estimation," accessed on 28.09.2021. [Online]. Available: https://www.v7labs.com/blog/human-pose-estimation-guide 2

[44] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1653–1660. 2

[45] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019. 2

[46] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *arXiv preprint arXiv:2006.10204*, 2020. 2

[47] "How to dress avant-garde," accessed on 28.09.2021. [Online]. Available: https://barbaraigongini.com/universe/blog/how-to-dress-avant-garde-what-goes-into-an-outfit/ 2

[48] "Mediapipe," accessed on 28.09.2021. [Online]. Available: https://google.github.io/mediapipe/ 2