Fashion Outfit Recommendation and Transformation Using Generative Adversarial Network Methods

Xinyi Meng

Department of Applied Mathematics, University of Washington, Seattle, USA

xmeng2@uw.edu

Abstract. Over the last few decades, generative adversarial network (GAN) methods have been increasingly applied and improved in fashion-related scenes, especially during the pandemic. This article provides an overview of several mainstream GAN models for fashion recommendation and fashion transformation, including compatibility matrix-regularized GAN, Outfit GAN, transformer-based GAN, conditional analogy GAN, disentangled cycle-consistency try-on network, and flow-navigated warping GAN. Within each GAN method, the improvements made to perform more related and deliciated are presented with detailed loss functions, the specific evaluation matrixes, and the examinations of generators or discriminators. Moreover, possible directions of improvements for future research and problems with existing models are involved. As for the common restrictions of included methods, including the problem of incomprehensive and non-diverse datasets, the standard evaluation of modish aesthetic value, and the number of fashion items involved in one outfit, are analyzed within fashion outfit recommendation and fashion transformation. Therefore, such restrictions provide possible directions for future analysis in fashion-related networks.

Keywords: Fashion recommendation, fashion transformation, generative adversarial networks.

1. Introduction

With the continuous development of COVID-19, online shopping and remote working have become critical parts of people's daily life, especially for those involved in the fashion industry. This paper focuses on how deep learning, the generative adversarial network methods, plays an essential role in the work and life of fashion workers and fashion lovers.

The generative adversarial network (GAN) structure can be considered the secondary field of machine learning, which comes from the inspiration of brain structure and its functionality [1]. The GANs model plays a crucial part in the 2D and 3D image generation and identification, speech synthesis, and text mining applications. These general applications are divided into different branches in the fashion industry, such as fashion outfit recommendation, personalized fashion design, fashion-oriented dialogue systems, and try-on [2]. According to the fact, the traditional fashion industry contains many images. It is needed to prove the quality of the pictures. GANs model is a significant application in this industry.

Machine learning, the GANs, changed the way how people would shop online in some sense. With the increased dataset and the detailed GANs networks, the fashion recommendation program stimulates the desire to shop online [3], particularly during the pandemic. However, purchasing online provides customers with large quantities of fashion items, leading customers finding hard to search for the ideal product. Meanwhile, unlike the commercial street shopping process, which includes physical try-on steps after finding the desired product, online shopping lacks customer-consultant communication and feedback, making the customer experience less satisfactory than offline shopping. Therefore, an online medium linking consumers and fashion producers needs to be proposed to fill this gap. Considering the characteristics of GANs, several GAN-based models, such as architecture-based GANs and objective function-based GANs, are presented for fashion recommendation and transformation [4].

As for the fashion outfit recommendation, it provide solutions for the people who are not sensitive to the esthetics. The fashion outfit recommendation is an application of GAN methods, which directly use different GANs methods for cloth match rules of high semantics. As for the online try-on process

is another crucial application for GAN methods in the fashion industry. With GANs network, it gives customers a way to upload their pictures and see how they would look with the fashion product suggested by the previous fashion recommendation process. With the fashion recommendation and transformation applications, the GANs network could dramatically increase the fashion industry's scale, which makes the fashion e-commerce still maintain a sure growth in the epidemic.

The remainder of this article is organized as follows: In section II, three GAN methods of fashion recommendation applications are introduced, which are compatibility matrix-regularized generative adversarial network (CMRGAN), outfitGAN, and transformer-based GAN. In section III, three GAN methods of fashion transformation applications are introduced, which are conditional analogy generative adversarial network (CAGAN), disentangled cycle-consistency try-on network (DCTON), and flow-navigated warping generative adversarial network (FW-GAN). At the end of section two and section three, the comparison and conclusion of each GAN fashion application are involved. Finally, concluding remarks are contained in section IV.

2. Fashion Outfit Recommendation Using Generative Adversarial Network Methods

2.1. Compatibility Matrix-Regularized Generative Adversarial Network (CMRGAN)

In [5], the authors showed how CMRGAN performed in fashion recommendation. In their work, they used a multi-model embedding method that took both visual and texture information as input data for each clothing. Such multi-model methods outperformed the original dropout layers [6], especially for large datasets. For the semantic compatibility space in cross-domain images, multi-model embedding method is used to transform individual image into a latent vector.

As for the loss functions, four loss functions are involved in the total loss function. The first is compatibility loss (L_{cross}). The paper established a triple set for the input image, consisting of compatible and incompatible items. The second is the feature matching loss (L_{feat}). This loss demonstrates the difference between the transformed feature and the proud-truth feature. The third is the traditional L1 loss (L_{L1}). Such loss ensures the generated image is close to the ground-truth image as much as possible. Finally, the adversarial loss (L_{adv}^{G}). As for the discriminator, the translated output and the sourer domain image are treated as fake pairs, and the target ground-truth image and source domain image are treated as real pairs. Like the traditional GAN structure, the discriminator tries to distinguish the actual parties and fake pairs. The alternative training process are provided in the generator, discriminator and the encoder in order to making the fake images are indistinguishable from the real images. In the end, the linear combination of four losses is introduced as the generator loss, which plays a min-max game. The total loss is:

$$\min_{E_{img}, E_{text}, G} \max_{D} L_{adv}^{G} + \lambda_1 L_{L1} + \lambda_2 L_{feat} + \lambda_3 L_{cross}, \quad (1)$$

The $\lambda_1, \lambda_2, \lambda_3$ are weights used to balance the loss objects.

As for the training part, the authors collected the fashion dataset from polypore.com. To evaluate the generated clothing pairs, the paper used several statistics in the dataset to represent the matching degree. The first score is the “mode score”, which is used to evaluate the quality of generated image. The “mode score” is the improvement of the inception score. However, the fashion dataset in [5] does not correlate with the dataset used for the training inception score. Therefore, the model does not need to use the original way to calculate the model score. The second score is MAE, the mean absolute error between the matching degree of compatible input pairs and the generated image pairs. The third statics score represents the matching score of the generated image pairs. This matrix shows that the ground-truth items is less compatible with the generated image. The fourth statics score is multi-scale structural similarity (MS-SSIM), a multi-scale variance of perpetual similarity metric. In particular, the CMRGAN model tries to get lower MS-SSIM scores since such a small score shows high image diversity. The last statics score is the deep attentional multimodal similarity (DAMS) score, which is

used to evaluate the matching degree of text-image since multi-model embedding is involved in the dataset—representing the sentence loss and defining the word loss.

Table 1. Result of CMRGAN method

Model	Mode Score↑	MAE↓	$S_f \geq S_r \uparrow$	MS-SSIM↓	DAMS	
					$L^S \downarrow$	$L^W \downarrow$
CMRGAN	2.3135	0.1076	911/1081	0.5205	6.32	9.64
Pix2Pix	2.3857	0.1	906/1081	0.6053	6.58	9.86
CycleGAN	2.289	-	-	0.4313	-	-
DiscoGAN	2.1744	-	-	0.5963	-	-

The results are shown in Figure 1. The authors compared the training result of CMRGAN with other state-of-art image translation models, which are pix2pix (a classic supervised model), CycleGAN (a traditional unsupervised model), and DiscoGAN (a GAN network that generates images of products in domain B given an image in domain A). In comparison, CMRGAN achieves similar realistic results since the model score of CMRGAN is close to each other. As for the MAE score, the CMRGAN model is close to the pix2pix. However, the CMRGAN model outperformed pix2pix in the score. Since the CMRGAN model aims to generate cloth pairs that at worse to have the same collocated score of input images, even though pix2pix performs well on generated image quality, it does poorly on the generated image creativity, which is the score. This result can also be observed in comparison with the MS-SSIM score. The CMRGAN model significantly drops on MS-SSIM, showing that the pix2pix model lacks diversity. Since both are decreased, the last evaluation score, DAMS, represents the model has a high text-image matching degree.

2.2. Outfit GAN

Using deep learning for fashion recommendation benefits fashion e-commerce, which speeds up product updates and solves the problem of a large quantity of overproduced items. For this purpose, Maryam Moosaei et al. propose the OutfitGAN, which generates a new fashion item to match the existing fashion items aesthetically. In this way, the fashion industry can arrange factory productivity by fashion item sales to achieve the purpose of "fashion-on-demand''.

As for the structure of OutfitGAN, which is different from traditional GAN, which has only one generator and one discriminator, Maryam Moosaei [7] et al. proposed the GAN with three stages of generator and discriminator pairs. Following generators receive previous generator's outputs and produce a higher resolution image. Therefore, the loss function should not only consider the real and fake images ($\mathcal{L}_{real/fake}$). In addition, to consider the minimax loss, the second factor that needs to be considered is whether the federated image belongs to the conditioned category ($\mathcal{L}_{cat}$), which comes to the softmax cross-entropy loss.

The next facet is whether the generated image is compatible with other items in the partial outfit ($\mathcal{L}_{comp}$). Finally, the last thing that needs to consider for the generator loss equation is whether the developed item has a similar distribution to the actual outfit ($\mathcal{L}_{KL}$). After putting all these considerations together, the loss function of each generator ($\mathcal{L}_{G_i}$) can be written as the linear combination of losses mentioned above:

$$\mathcal{L}_{G_i} = \alpha\mathcal{L}_{real/fake} + \beta\mathcal{L}_{cat} + \lambda_i\mathcal{L}_{comp} + \omega\mathcal{L}_{KL}, \quad (2)$$

The $\alpha, \beta, \lambda_i, \omega$ are weights used to balance the loss objects.

As for the discriminator, the loss function also needs to regard the ability to distinguish real and fake images and the ability that the generated image is in the correct category as the evaluation. The loss of each discriminate is also the linear combination of the above losses.

Table 2. Result of OutfitGAN method

Dataset	Items	Quality (IS)	Diversity (MS-SSIM)	Compatibility (m_s)
Polyvore	original	1.125	0.146	0.900
	generated	1.117	0.153	0.859
iFashion	original	1.128	0.151	0.910
	generated	1.119	0.180	0.863

After training such a deep learning model with Polyvore and iFashion dataset (as shown in Figure 2), two statics scores are involved in the model evolution: the first is quality and diversity, which is represented by the inception score (IS), and the multi-scale structural similarity metric (MS-SSIM), and the compatibility score. From the experiment result, we can see that even though the quality of OutfitGAN does not perform better than the original one, which is a slight decrease, the diversity of OutfitGAN does well perform, especially in the iFashion dataset. In addition, one exciting aspect of OutfitGAN is that the specific number of items is not the requirement of OutfitGAN, which is a dramatic improvement in fashion recommendations.

2.3. Transformer-based GAN

The [7] shows that, in some sense, it solves the problem that GAN-based methods mainly generate a fixed number of items. However, if the number of items changes, the whole model needs to be retrained. The second problem is that the existing model process the visual and textural information separately, or even only extract the visual features. The attributes of the input items and the evaluation's reliability of the characteristics of the received items are not guaranteed by such method. Therefore, the third deep learning model: multimodal transformer-based GAN with cross-model attention, is designed to solve these problems [2].

The first change Valery Volokha and Klavdiya Bochenina perform is cross-model attention (CMA) and self-attention, raised by Ashish Vaswani et al. [8]. The query, key, and value are performed as the representatives in each sequence. In this way, the multimodal transformer block can explore two modalities, visual and textual, and extend the scheme.

The second remarkable change in the paper is performing multimodal transformer-based GAN, which is involved in the generator (as shown in Figure 3). In the transformer-based GAN model, the compatible no-noise sequence are translated from the given items aligned between modalities by the generator.

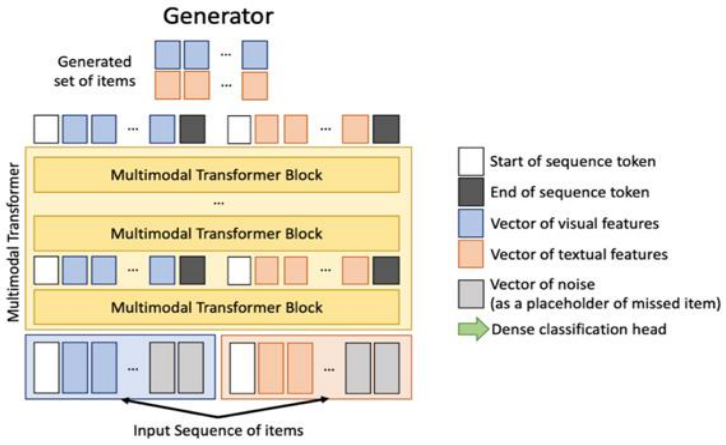


Figure 1. The base of Generator [8]

The main structure of the discriminator (as shown in Figure 1) is similar to the generator but with several changes. In transformer-based GAN structure, their model is required to generate realistic vectors and to learn the compatibility of input vectors and generated new compatible vectors. For such purpose, the team presents a compatibility discriminator for compatibility modeling.

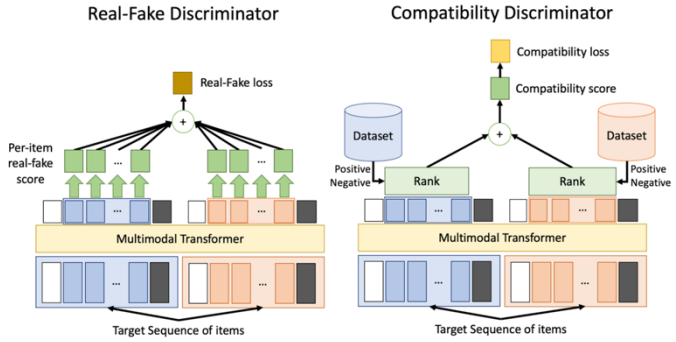


Figure 2. The base of Real-Fake Discriminator and the base of Compatibility Discriminator [8]

After training the transformer-based GAN model with FOTOS, similar to Polyvore in dataset size, the paper presented the model result by comparing the statics score with other baseline methods such as FHN, Bi-LSTM-VSE, NGNN, HFGN, and MGCM. The result is evaluated as the fill-in-the-blank task. Such a task is presented as a recommendation by randomly removing fashion items from an outfit, and the model must select the correct answers to fill in the blank.

Table 3. Result of Transformer-based GAN

Method	FLTB (2 items)	FLTB (3 items)	FLTB (4 items)
FHN	0.697	0.669	0.669
Bi-LSTM-VSE	0.724	0.776	0.753
NGNN	0.791	0.765	0.765
HFGN	0.826	0.801	0.800
MGCM	0.822	0.817	0.829
Transformer-based GAN	0.878	0.863	0.911

The result shows that the transformer-based GAN model performs well in such FLTB tasks, representing outperforming tasks for all baseline models, which reached 0.878 accuracies in filling one blacked item by one of four candidates by one input item.

2.4. Possible Improvements

From the above fashion recommendation methods, there are several improvements made by those models. One improvement is the semantic alignment in the generated items. Also, the improvement made by transformer-based GAN is surprising.

However, other things still need to be considered. Improving the quality and diversity of generated cloth is the end goal for most fashion recommendation modules. Although all models involved in previous sections show quality improvements, the continued pursuit of higher quality generated images leads the researchers to create or improve the models.

In actual circumstances, one outfit could contain several fashion items such as headbands, necklaces, etc. Currently, such fashion recommendation tasks mostly rely on the human fashion consultant. Therefore, one incoming question is could the model recommend not only clothing but also accessories. In other words, could the generated fashion image contain more than one fashion item? This possible improvement is worth the consideration of further research.

The other one is the evaluation of fashion recommendations. The above fashion recommendation models and the dataset does not have the specific assessment authority, or matching degree, of the recommended fashion items according to aesthetics. Therefore, this crucial evaluation standard is worth adding to the results table for model evaluation.

Besides the standard fashion recommendation evaluation, one problem that needs to be considered is the information involved in the datasets. The category within each group does not evenly distribute. For example, in the iFashion dataset, the “Gender” group labeled 92.3% of images of the dataset, but the “Pattern” group only tagged 30.8% of images of the dataset [9]. These uneven labels create non-accuracy in the learning process. Furthermore, some datasets are pre-trained, showing that the results of the pre-trained datasets could vastly outperform other datasets. Sheng Guo et al. showed that some

datasets, such as Clothes-50K and DeepFashion, would be better target training datasets for machine learning than others.

The fashion recommendation not only needs to consider the matching degree of clothing items but also the popularity of the clothes. With all the fashion recommendation methods, one actual circumstance lacking in these models is the fashion trends in different seasons, countries, age groups, etc. From the perspective of simplifying the model and speeding up the training time, such a job should be performed by different datasets that involve a diversity of fashion preferences. Fortunately, with the development of online datasets, large internet companies can provide or mesmerize the personal preference of fashion items, making the fashion recommendation dataset available for training the particular person's model [10].

Finally, the end goal of all machine learning is to develop a model that requires less time. With this directional goal, the multi-model method, such as CMRGAN, requires training time improvements [11]. With such training cost, choosing the proper model optimization requires further experiments.

3. Fashion Transformation Using Generative Adversarial Network Methods

3.1. Conditional Analogy Generative Adversarial Network (CAGAN)

The fashion transformation, or the try-on application, is representative of speed in fashion e-commerce. Such a try-on application shows a better solution for the fashion industry to reduce the input funding on making photographs with professional models and avoid significant consumption in modeling humans and fashion items in 3D objects with 2D image collections. The CAGAN provides better deep learning neural network for the fashion industry.

The primary purpose of the CAGAN model is to train an image-to-image translation model that changes a specific clothing item (y_i) with another clothing item (y_j) on a given human (x_i). The generator in the CAGAN model is similar to the original GAN model, which creates fake data by receiving feedback from the discriminator. However, this model's discriminator is different from the original one. Nikolay Jetchev and Urs Bergmann proposed that the discriminator needs to solve two problems: does the human image look reasonable, and does the fashion item looks well painted on the human model? The second problem is much harder to solve since the data in [12] only contains a few examples of specific humans. The second criterion cannot be entirely modeled with a Euclidean loss. Therefore, here comes the linear combination of generator loss and discriminator loss:

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \gamma_i \mathcal{L}_{id}(G) + \gamma_c \mathcal{L}_{cyc}(G), \tag{3}$$

Compare the CAGAN loss equation with the classical GAN loss function, which learns to distinguish the ground-true data from generated data. The adversarial loss term defines another negative: the human model and new clothing item, different from the original one, do not appear on the given human. Besides another part of the equation, the term is also worth noticing. Such a time prevents the irrelevant model parts for swapped clothing items. Finally, the loss is designed to make the generator more stable by leaving the other parts of the image which are irrelevant for training unchanged.

The data, which contains 15000 images that the paper used, is given by Zalando SE, one of the biggest e-commerce fashion companies. After training with such a dataset, we can see results from randomly chosen pictures that the model has the desired behavior: swapping clothing items without changing other irrelevant parts.

3.2. Disentangled Cycle-consistency Try-on Network (DCTON)

According to the previous model's result, the CAGAN is a well-performed model of fashion transformation. However, taking a closer look at the CAGAN model's results, one can find that the picture generated for try-on does not provide sufficient clarity. In other words, the CAGAN model

can still not simulate the texture of clothes, human skin, and the background (non-clothing simulation). In this way, the model is insufficient for high-quality try-on. Therefore, the new fashion transformation model [13] was raised by Chongjian Ge et al.: the disentangled cycle-consistency try-on network (DCTON). The DCTON model is improved basically in three parts. The first one is generating more clothing item details such as logo. Besides, the second improvement is the skin simulation for the included parts, Finally, improvement is the image position model for output generation.

Similarly, several losses are introduced for discriminators and generators. The similarity of appearance distributions between the try-on images and the target images is guaranteed by two discriminators, presenting in the adversarial loss. The synthesized try-on results and the corresponding skin regions are guaranteed by the consistency loss. Besides, as for the contents that included skin parts and clothing items, the model presents the content preserving loss to achieve such purpose. Finally, the similarity of CNN feature appearance in the warped clothes is guaranteed by the perceptual loss. Also, the final objective function is the linear combination of all mentioned loss expressions.

Table 4. Result of DCTON

Methods	Dataset	IS↑	SSIM↑	FID↓
CAGAN [12]	VITON	2.56±0.09	0.74	47.34
VITON [14]	VITON	2.29±0.07	0.74	55.71
ACGPN [15]	VITON	2.69±0.12	0.81	16.64
DCTON [13]	VITON-HD	2.84±0.10	0.81	15.55

Regarding the implementations, the team involves the VITON and VITON-HD datasets for the training model. There are three statics results involved in the evaluation, which are fréchet inception distance (FID), structural similarity (SSIM), and inception score (IS). As shown in the table above, the IS score of the DCTON model is over performs other methods dramatically, and the SSIM score surpasses others. Besides, the DCTON method is well performed in the quality of synthesized images since the DCTON model shows a lower FID score.

Table 5. Time and computational cost of DCGAN

Methods	Dataset	Training Time	#Params	FLOPS	FPS
ACGPN [15]	VITON	~40h	139M	206G	10
DCTON [13]	VITON	~44h	153M	194G	19

Besides the high-quality of visual performance, the DCTON model proposed by Chongjiang Ge and his team shows that such a model uses fewer computational resources. The paper presents the computational costs of ACPGN and DCTON, indicating that with the same dataset and GPU, the DCTON uses about the same time as ACPGN but trains more parameters, which is a tremendous improvement of the fashion transformation model.

3.3. Flow-navigated Warping Generative Adversarial Network (FW-GAN)

Previous methods focused on the 2D (pictures) try-on deep learning method for fashion transformation. A reasonable improvement came into researchers' minds: could video try-on be possible? To solve this problem, Haoye Dong et al. moved further with a video visual try-on deep learning method: flow-navigated warping generative adversarial network (FW-GAN) [16]. In other words, using a person's image, a given clothing item, and a series of required poses, the model tries to synthesize a realistic-looking video presenting the person and the clothing item. However, one obvious problem came to the surface: with different poses, different parts of clothing items could be changed, including texture, color, etc. Therefore, the FW-GAN model is designed to solve this problem.

In detail, three main components are involved in the FW-GAN model. First, the high quality and smooth generated video is enforced by a flow-navigated module. Second, a warping net is adapted to evaluate the transformation parameters grid which warps the desired clothes. Finally, a figure parsing

constraint loss is introduced to restrict human pose to ensure the overall consistency. For the generator, the paper proposes a conditional GAN suture for the generator to memories the conditional probability distribution.

Similarly, the generator is optimized throng the standard minimax optimization. As for the discriminator, several papers report that using multiple discriminators may reduce the possibility of collapse in GAN training. However, with visual quality and temporal consistency requirements, two discriminators are involved in the model structure: a frame discriminator and a flow-embedding discriminator. The frame discriminator is designed for each frame's visual quality, and the flow-embedding discrimination is designed for the neighboring structures' temporal consistence, which helps the video more smoothly. Besides, five losses are introduced in the objective function. The perceptual loss, which contains two different features from the pre-trained VGG network and discriminator, the parsing constraint loss, which captures the local information of sub-parts, the flow loss adopted from FlowNet, and finally, more warped items' pixel are learned through the generator by grid loss.

Table 6. Result of FW-GAN

Fréchet Inception Distance	I3D	ResNeXt - 101
CP-VTON [17]	32.35	159.50
VITON [14]	30.05	129.74
FW-GAN [16]	7.052	23.94

For the implementation, a video dataset video virtual try-on dataset (VVT), is used for training the model. With the statics evaluation of FID, I3D, and 3D-ResNeXt-101, which is used to extract the spatial-temporal feature vectors, the table shows that the FW-GAN is capable of synthesizing the desired image and the selected video and dramatically outperformed the Baseline method CP-VTON and GAN-VTON methods. Two exciting things worth paying attention to are the critical role of the flow model and the grid model. Without a grid model, the network leads to low-resolution video, and clothing items' patterns are lost. Without a flow module, the discriminator fails to obtain the spatiotemporal smoothing. In conclusion, adding these two structures makes the FW-GAN outperforms the baseline models.

3.4. Possible Improvements

Those papers show incredible improvements in fashion transformation application from the above method, such as high-quality generated pictures (particularly on logo) and video try-on models. However, several shortages are also presented in the current research.

One obvious problem is the background and pose interference on try-on applications [20]. The datasets used by most models are uninformed black or white backgrounds [12]. The people's pose is also a problem for current GAN models. However, in the actual scenario, people could try on the fashion items in different environments, which means the environment is impossible to be that uniform. This would dramatically influence the outcome of generating the image. Therefore, more improvements need to be made for in-uniformed background try-on models.

The CAGAN model in [12] can also be improved using other datasets. The dataset in [12] uses RGB rather than Lab, which is one possible improvement of trying other color spaces. Besides, the dataset used does not contain textural information, which creates inaccuracy in the fine-grained classification. Since the advantage of the conditional generative model is being able to direct the image generation, as discussed in [18], without such textural information, the conditional generative model can not take advantage of it.

With the dataset that contains textural information, the conditional generative adversarial model can take advantage of such class labels to generate further detailed sub-categories for fashion recommendation.

Other than the datasets problems, the limitation of the evaluation matrix also needs to be considered. The evaluation matrix involved in [13] is valuable to consider. However, some of those

evaluation matrices still contain drawbacks [19]. For example, the inception score could also perform well even when the model collapses, which easily happens in GANs model. Besides, the fréchet inception distance could also perform well even in the overfitting problem. Therefore, these limitations should be involved in the consideration.

Also, as mentioned before, the same goal of all machine learning – building a model which takes less time. However, most papers on the fashion transformation GAN model focus on the results of generated pictures and statics results of the trained model. Only a few reports contain the training time results. Therefore, the training time is also considering improvements in fashion transformation applications.

4. Conclusion

In this paper, three GAN-based models are involved both in fashion recommendation and fashion transformation. Those models show incredible matching items and generated pictures with improvements in loss functions, generator structure, and discriminator structure. All models outperform the ground-based models. With several datasets, such as Polyvore and iFashion dataset, all models show convincing results.

By implementing those models, several possible improvements are needed for further development. The first one is the datasets. The datasets' information must be considered before training all those models. Secondly, the problem of having several fashion items in an outfit and the problem of quality is the evaluation of fashion recommendations and aesthetic value. This kind of problem currently cannot be solved or answered by the model proposed by these papers. Therefore, further research or standards must be involved in fashion-related models. Third, the standard or largely agreed evaluation method (or matrix) is involved in the fashion recommendation and fashion transformation. Finally, the issue that has been discussed for a long time is the model's efficiency.

References

[1] A. Aggarwal, M. Mittal and G. Battineni, “Generative adversarial network: An overview of theory and applications”, Processing of the IJIM Data Insights, vol. 1, no. 1, 2021.

[2] V. Volokha and K. Bochenina, “Content-Aware Generative Model for Multi-Item Outfit Recommendation”, No. 8016. EasyChair, 2022.

[3] S. Kumar, and M.Das Gupta, “GAN: complementary Fasihon Item Recommendation”, Processing of ArXiv, abs: 1906.05596.

[4] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan and Y. Zheng, "Recent Progress on Generative Adversarial Networks (GANs): A Survey," in IEEE Access, vol. 7, pp. 36322-36333, 2019.

[5] L. Liu, H. Zhang, and D. Zhou, “Clothing generation by multi-modal embedding: A compatibility matrix-regularized GAN model”, Image and Vision Computing, Vol 107, 2021.

[6] Z. Shen, Z. He, and X. Xue, “MEAL: Multimodel Ensembel Via Adversarial Learnning,” Processing of the AAAI cnference om Artificial Intelligence, Vol. 33, no. 01, 2019.

[7] M. Moosaei, Y. Lin, A. Akhazhanov, H. Chen, F. Wang, and H. Yang, “OutfitGAN: Learning Compatible Items for Generative Fashion Outfits” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2022, pp. 2273-2277.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need”, Processing of Advanced in Neural Information Processing Systems 30, 2017.

[9] S. Guo, W. Huang, X. Zhang, P. Srikhanta, Y. Cui, Y. Li, H. Adam, M. R. Scott, and S. Belongie, “The iMaterialist Fashion Attribute Dataset” Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 0-0.

[10] Y. Deldjoo, F. Nazary, A. Ramisa, J. Mcauley, G. Oellegrini, A. Bellogin, and T. D. Noia, Processing of ArXiv, abs: 2202.02757.

[11] H. Wu, D. An, X. Zhu, Z. Zhang, G. Fan, and Z. Hua, “Muti-source material image optimized selection based multi-option composition”, Image and Vision Computing, Vol 107, 2021.

[12] N. Jetchev, and U. Bergmann, “The Conditional Analogy GAN: Swapping Fashion Articles on People Images”, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2287-2292.

[13] C. Ge, Y. Song, Y. Ge, H. Yang, W. Liu, and P. Luo, “Disentangled Cycle Consistency for Highly-Realistic Virtual Try-On”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 16928-16937.

[14] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S Davis, “Viton: An image-based virtual try-on network”, Proceeding of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[15] H. Yang, R. Zhang, X. Guo, W. Liu, W. Zuo, and P. Luo, “Towards photo-realistic virtual try-on by adaptively generating-preserving image content”, Processing of the IEEE International Conference on Computer Vision (ICCV), 2020.

[16] H. Dong, X. Liang, X. Shen, B. Wu, B. Chen, and J. Yin, “FW-GAN: Flow-Navigated Warping GAN for Video Virtual Try-On”, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1161-1170.

[17] B. Wang, H. Zhang, X. Liang, Y. Chen, and L. L, “Toward characteristic-preserving image-based virtual try-on network”, Processing of ArXiv, abs: 1807.07688.

[18] M. Mirza, and S. Osindero, “Conditional Generative Adversarial Nets”, Processing of ArXiv, abs: 1411.1784.

[19] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Rencent Progress on Generative Adversarial Networks (GANs): A survey", in IEEE Access, vol. 7, pp. 36322-36333, 2019.

[20] M. F. Hashmi, B. K. K. Ashish, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "FashionFit: Analysis of Mapping 3D Pose and Neural Body Fit for Custom Virtual Try-On," in IEEE Access, vol. 8, pp. 91603-91615, 2020.