Intelligent Fashion Generation System

Kavya Khatter
Dept of CSE
PES University
kavyakhatter1999@gmail.com

Daksha Singhal
Dept of CSE
PES University
daksha.singhal@gmail.com

Adithya M Vardhan
Dept of CSE
PES University
adivar1999@gmail.com

Srinivas KS
Dept of CSE
PES University
srinivasks@pes.edu

***Abstract*—The fashion industry has always been one of the most lucrative markets in terms of products and design. Every designer is always looking for new ways to come up with ideas and inspirations. Every new design and concept could be a new trend in fashion. In this paper, we discuss the applications of Generative Models to train and develop a vast array of fashion apparel, given an adequate dataset. Characteristics such as texture, shape, design and material are some of the parameters that we take into consideration while creating variations in the output. Features, specific to a designer, can be achieved by loading the previous works of the designer into the model. We will cover the various methodologies used to achieve the output and the best solution that we found. Such applications would find an interesting use case in designer boutiques such as H&M and ZARA as well as existing online boutiques such as SHEIN.**

Keywords: Fashion, Generative Adversarial Networks, Deep Learning

There are broadly VI sections in this paper consisting of an introduction, literature survey, Implementation, Observation and Results, Conclusion and Future Works and Acknowledgement.

I. INTRODUCTION

In the fashion industry, every designer has a unique taste, a unique sense of style and a designated aesthetic to their products. To deliver such a large variety of designs, fashion designers spend months together in putting up a montage suited to a particular theme in the time and place and are compensated accordingly. But what if we could cut short that process by automating the process? Study the previous designs of the designer and produce a large variety of apparel that can be selected to help push the process of elimination.

Using variations of generator models, we aim to develop a model that takes in several inputs varying in colour, shape, design and texture, then used to create randomized outputs of different textures and colours which bear a similar resemblance to the original works of the designer but also new combinations the designer could contemplate in order to gain a better insight.

To implement this, we propose to use an improvised version of GANs to provide an unsupervised separation of high-level attributes of the generated images. The generated images have been known to suffer from low-resolution, a problem to which we find a solution in the application of Super-Resolution GANs.

The application of such a product lies in many fields within fashion. Besides the proposed use case in apparel, implementations lie in designing of large draperies, accessories and jewellery.

II. LITERATURE SURVEY

1) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network

This paper offers a solution to the low-resolution images created from the generator model. By using GANs, the solution effectively increases the resolution of the photo by up to 4x times while creating photo-realistic natural images.

2) Generate high-resolution images with Generative Auto-Encoder

An innovative solution of using a hybrid Variational Auto-Encoder model along with the Generative Adversarial Network to create high-resolution images. This paper inputs the results of a generator as well as an encoder into the discriminator to identify the authenticity. This bypasses the pitfalls that GANs and VAEs encounter while running by themselves.

3) A Style-Based Generator Architecture for Generative Adversarial Networks

This paper proposes an unsupervised separation of high-level attributes and stochastic variation in the generated images and enables intuitive, scale-specific control of the synthesis. It also introduces a new generator leading to better interpolation properties and also disentangles the latent factors of variation better and therefore introduces two new methods.

4) DeepWear: a Case Study of Collaborative Design between Human and Artificial Intelligence

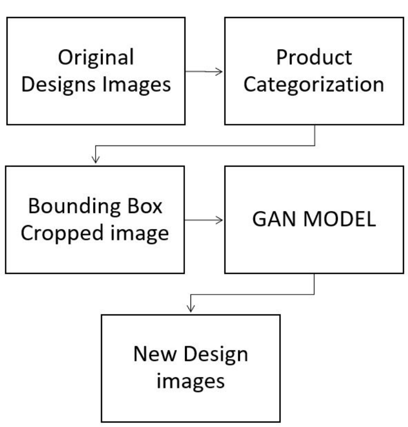
DeepWear is a practical designing clothes system to generate images using deep convolutional generative adversarial networks and designers make clothes by receiving instruction from those images. The system was then evaluated by comparing the credibility of actually sold clothes with the generated clothes.

5) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

A DCGAN is an extension of vanilla GAN, which uses convolutional and convolutional-transpose layers in the discriminator and generator, respectively. The generator consists of Deconvolution layers or more accurately, Transposed Convolution layers that basically perform the reverse of a Convolution operation while the discriminator is a binary classifier that consists of several convolution layers.

Prior works in improving the quality of the image have been proposed in [][]. Our work revolves around the same idea i.e first generating images using Generative Adversarial Networks followed by the refinements to the image generated in the first stage.

III.IMPLEMENTATION



a. Dataset

Since there are a few processing steps before actual training of the network for the generation of new designs, we have used an image dataset with their corresponding coordinates of the apparel needed for the bounding box technique. For the training step, we have used Kaggle product dataset which consisted of 44000 images of various categories and also scrapped 10,000+ images from different e-commerce websites.

b. System Framework

This section presents the implementation of details of the different approaches to build the fashion generation system. There are broadly two stages involved in the generation of new designs.

Data preprocessing

Classify the images into predefined categories

Given input images of different fashion styles, we have built the model with 3 layers (1 layer to flatten the image to a 28 x 28 = 784 vector, 1 layer with 128 neurons and Relu activation function & 1 layer with 10 neurons and the Softmax function). We have 10 Neurons because we have 10 labels for the image data set.

Table 1: Label for Clothing Classification

T-shirt	trouser	pullover	dress	coat
Sandal	Shirt	sneaker	bag	saree

Bounding Box approach

An object detection model was built using YOLO architecture that takes real-world images and detects and classifies apparel. Thus, using this model we detected the apparels in the images and cropped the specific clothing item and stored in the database.

Modelling

This section describes the model incorporated to generate high-resolution apparel images from the preprocessed dataset.The underlying architecture for the generation of fashion images is generative adversarial network (GAN) architecture. There are primarily 2 parts of generative adversarial network model :

- The generator is a function that learns to generate synthetic data from random data in terms of a probabilistic model.

- The discriminator takes the generated image from the generator as negative samples and then learns to differentiate the generator's fake image from the actual image. The discriminator penalizes the generator for producing implausible results.

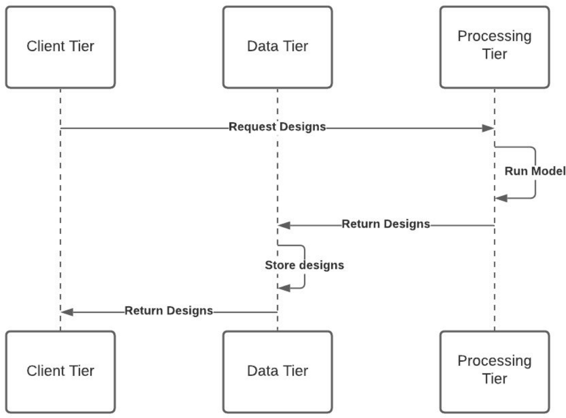


Fig:1 Interface Diagram of Generating Designs

DCGANs + SRGANS

DCGANS

A DCGAN is an annexe of the vanilla GAN besides its use of convolutional layers in the discriminator and convolutional-transpose layers in the generator. Three major changes are made in the DCGAN architecture that is different from CNN architecture.

First, Deterministic spatial pooling functions with stridden convolutions are replaced by All Convolutional Net letting the generator to acquire its own spatial downsampling.

Second, the fully connected layers such as the global average pooling are eliminated that is on top of convolutional features as it reduces the convergence speed greatly. Concurrently, the foremost GAN layer receives as input which is a uniform noise distribution and outputs a four-dimensional tensor to be used as the start of the convolution stack. In the case of the discriminator, the final convolution layer is flattened and fed into a single sigmoid yield. Finally, both the discriminator and the generator undergo batch normalization which steadies learning by normalizing the contribution to every unit to have zero mean and unit difference and blocks the generator from deteriorating all examples to a solitary point.

Within the generator, most of the activation functions fall under ReLU activation excluding the output layer which uses a Tanh function. In the case of the discriminator, it uses Leaky ReLU activation function for all instances.

SRGANS

Generative Adversarial Networks are extremely strong structures to be used in developing images of high quality while also being perceptually satisfying. The model outlined in the paper uses the concept to form a perceptual loss function for Super-Resolution photo realistic. This model contains three important components to be discussed, the perceptual loss employed and the changes made to each of the Generator and Discriminator.

Two losses, the adversarial loss and the content loss, together form the perceptual loss function proposed by the paper, in the attempt to train the generator G into fooling the discriminator D which is trained to distinguish the real images from the generated ones.

The perceptual loss is calculated by computing the two parts of the loss relevant to maintain the perceptual clarity of the image, content loss and adversarial loss. Rather than relying on the popular loss at a pixel-level, the authors of this research define a loss based on the ReLu activation layers called VGG loss. The adversarial loss motivates the generator in approaches that help fool the discriminator, by resorting to images that border at the edge of the natural images.

At the core of the Generator, networks are a certain number (B) residual blocks of the identical layout. The representation of the architecture is presented in figure 1. Our approach outlines the use of two convolutional layers with 64 feature maps and miniature 3x3 kernels followed by a layer performing batch-normalization and finally, an activation layer with ParametricReLu as the activation function.

To distinguish between the Super-Resolution images generated by the generator G and the real high-Resolution images, we use the Discriminator model of the GAN architecture as shown in the figure. Similar to the generator model, we employ LeakyReLu activation but we circumvent max-pooling throughout the network. This model consists of 8 convolutional layers with varying number of third-degree filter kernels, each greater than the previous by a factor of 2, starting at 64 and working itself all the way upto 512 kernels as in the network specified previously as the VGG network. The 512 feature maps that arrive as output to this model are then followed by two dense layers and a sigmoid activation function. This promotes perceptually improved solutions found in the context of the real high-resolution images.

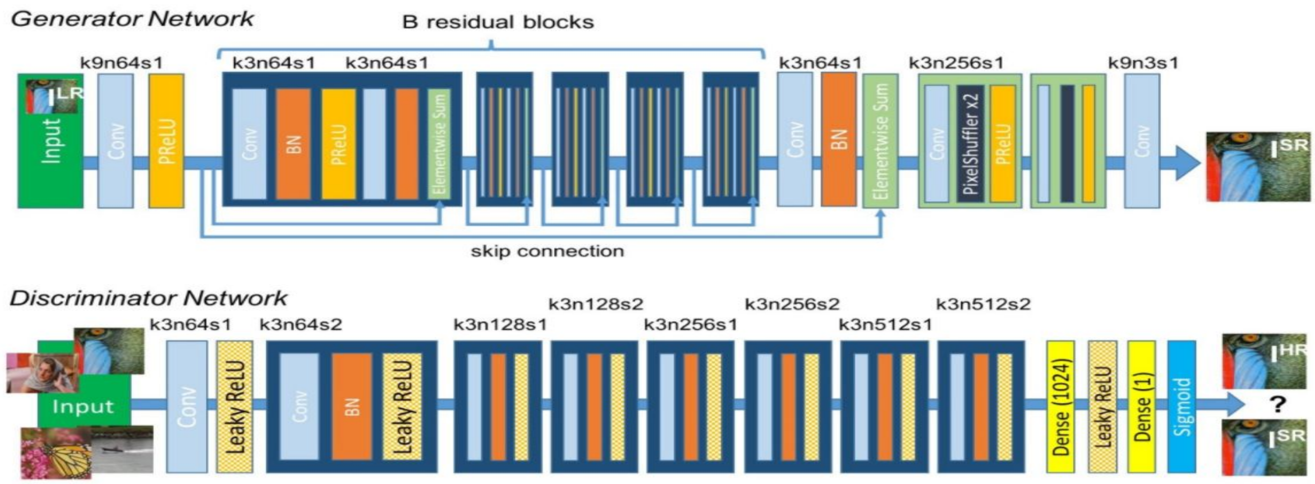


Fig:2 Model Architecture

IV. OBSERVATION AND RESULTS

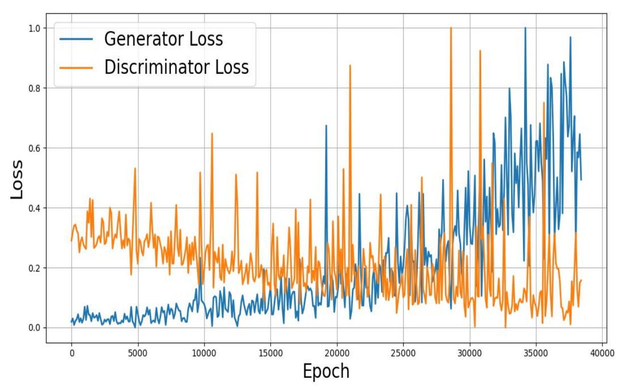


Fig:3 Model Architecture

The metric used for calculating the image quality for the generated images is BRISQUE score that calculates the image quality of an image with no reference using the Blind/Referenceless Image Spatial Quality Evaluator. The average BRISQUE score for our generated images is 34..

DC GAN model is trained on samples of 5K images of sarees and t-shirts for 50K epochs and a batch size of 32 .The optimizer used in the model is adam with the learning rate of 0.0004 .The training is implemented using PyTorch and ran for around 24 hours.
As observed from fig 3, the generator loss stabilizes at 20K epoch . The face is blurry and we therefore conclude that the DCGAN model focuses on the structure and design of the clothing item.

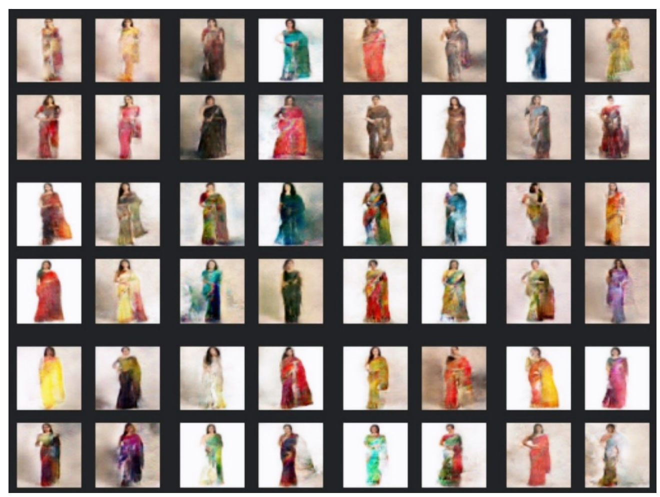


Fig:4 New Saree Designs

Fig:5 New T-shirt Designs

V.CONCLUSION AND FUTURE WORKS

Through this research,we have successfully built an application where the client can request for new designs by inputting his original design images. For generating new designs we have implemented DC-GANs which provided us with the desired results given the good quality adequate number of images.

For Future Works, we will be training our model on a higher resolution (512x512) using powerful GPUs.

VI. ACKNOWLEDGEMENT

We would like to thank our mentor and guide Prof. Srinivas KS for the support and guidance over the course of the project. Finally, we would like to thank the Computer Science Department of PES University for providing the opportunity to work on this project.

REFERENCES

[1] Natsumi Kato, Naoya Muramatsu, Hiroyuki Ozone, Yoichi Ochiai, Daitetsu Sato University of Tsukuba, *DeepWear: a Case Study of Collaborative Design between Human and Artificial Intelligence*

[2] Tero Karras, Samuli Laine, Timo Aila NVIDIA, *A Style-Based Generator Architecture for Generative Adversarial Networks*

[3] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi Twitter, *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*

[4] Roman Zeyde, Michael Elad, and Matan Protter, *On Single Image Scale-Up Using Sparse-Representations*

[5] Alec Radford, Luke Metz, and Soumith Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks* In International Conference on Learning Representations (ICLR), 2016.

[6] Seyedarian Hosseini, Thomas Boquet, Wojciech Stokowiec, Ying Zhang, Christian Jauvin, Chris Pal, Fashion-Gen: *The Generative Fashion Dataset and Challenge*

[7] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, *Variational autoencoder for deep learning of images, labels and captions*, in Advances in neural information processing systems, 2016, pp. 2352–2360.

[8] Yanghua Jin, Jiakai Zhang, Minjun Li, Yingtao Tian,Huachun Zhu, and Zhihao Fang. 2017. *Towards the Automatic Anime Characters Creation with Generative Adversarial Networks.* arXiv preprint arXiv:1708.05509 (2017).

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2016. *Image-to-image translation with conditional adversarial networks.* arXiv preprint arXiv:1611.07004 (2016).

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. *Generative adversarial nets. In Advances in neural information processing systems*. 2672–2680.

[11] Yang Hu, Xi Yi, and Larry S Davis. 2015. Collaborative fashion recommendation: *A functional tensor factorization approach.* In Proceedings of the 23rd ACM international conference on Multimedia. ACM, 129–138.

[12] Lixuan Yang, Helena Rodriguez, Michel Crucianu, and Marin Ferecatu. 2017. *Fully Convolutional Network with Superpixel Parsing for Fashion Web Image Segmentation*. In International Conference on Multimedia Modeling. Springer, 139–151.

[13] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. *Deepfashion: Powering robust clothes recognition and retrieval with rich annotations.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1096–1104.

[14] Zequn Li, Honglei Li, and Ling Shao. 2016. Improving Online *Customer Shopping Experience with Computer Vision and Machine Learning Methods.* In International Conference on HCI in Business, Government and Organizations. Springer, 427–436.

[15] Haosha Wang and Khaled Rasheed. 2014. *Artificial intelligence in clothing fashion.* In Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 1.