Fashion Fitting in Deep Learning

Hui Kwat Kong 20123133
Song Yangyang 20534320

1. Introduction

With the fast-paced lifestyle and rapid development of the internet technology, online shopping is more preferred than the traditional shopping in shop. However, it is difficult to predict how the clothes fits the person without seeing the whole image of wearing the particular clothes. Hence, this project analyzes the state-of-the-art model using DeepFashion dataset, pose transfer technologies to transfer arbitrary clothing to a person's clothing.

2. Data Description

DeepFashion [5] contains over 800,000 diverse fashion images ranging from well-posed shop images to unconstrained consumer photos. Each image in this dataset is labeled with 50 categories, 1000 descriptive attributes, bounding box and clothing landmarks. The dataset also contains over 300,000 cross-pose/cross-domain image pairs.

In-shop Clothes Retrieval Benchmark: It evaluates the performance of in-shop Clothes Retrieval. The dataset contains large pose and scale variations for each cloth.

3. Related work

3.1. M2E-Try On Net

The virtual Try-On network, M2E-Try On Net [6], transfers the clothes from a model image to a person image without the need of any clean product images, which is able to align the poses between the model and the target person with preserving the model's clothes.

3.2. SwapNet

SwapNet [1] is able to swap the clothing between a pair of images while preserving the pose and body shape. There are two stages including warping stage and texturing stage.

3.3. A Variational U-Net for Conditional Appearance and Shape Generation

U-Net [2] is used for mapping from shapes to target images and for conditioning potential representations of variational autoencoders about appearance.

4. Technical overview

4.1. Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep neural network for analyzing visual imagery, which typically consists of convolution layer, pooling layer and fully connected layer, with ReLU as the activation functions.

4.2. VGG16

VGG (Visual Geometry Group) 16 [4], a deep learning CNN model, utilizes 3x3 convolutional layer (instead of 5x5, 7x7, 11x11), which outperforms AlexNet.

4.3. Generative adversarial network

Generative Adversarial Network (GAN) [3] is a popular and cutting edge of unsupervised learning method with two neural networks (Generator and Discriminator Network) contesting with each other in a zero-sum game framework.

5. Expected results

Given the Model with clothes (M_c) and pose (M_p) and Target with clothes (T_c) and pose (T_p), the task is to transfer the Model's clothes with pose M(M_c, M_p) to an arbitrary Target person T(T_c, T_p) (The operation can be written as T(T_c, T_p) $\rightarrow$ T(M_c, T_p)). The result will be delivered in two stages, in other word, transferring the pose first and then the clothes.

5.1. First stage: Pose transferring

The first stage is to generate a model's clothes image with the target person pose, i.e. M(M_c, M_p) $\rightarrow$ M(M_c, T_p). There are two approaches:

1. Generate the "densepose" of the Model (M_d) and Target (T_d) by transfer learning. Then we learn how to concatenate the M_c, M_d, T_d to M(M_c, T_p).
2. Apply the state-of-the-art model Variational U-Net as transferring learning to perform the pose transfer.

5.2. Second stage: Texture transferring

After generated the transferred pose model image M(M_c, T_p), this stage is to transfer the clothing part:

$$T(T_c, T_p) \xrightarrow{M(M_c, T_p)} T(M_c, T_p) \quad (1)$$

The idea is to utilize GAN framework to discriminate our truth target $T_{truth}(T_c, T_p)$ and $T_{GAN}(T_c, T_p)$. The best situation is that we have T(M_c, T_p) and T(T_c, T_p) at the same time, but it is hard to get such pair clothes data. We perform data augmentation of $T_{truth}(T_c, T_p)$ to avoid the identical mapping. The network structure consists of GAN, Interest (ROI) pooling and the result from first stage.

References

[1] Amit Raj, Patsorn Sangkloy, Huiwen Chang, James Hays, Duygu Ceylan, Jingwan Lu. SwapNet: Image Based Garment Transfer. ECCV (12) 2018: 679-695 1

[2] Esser P, Sutter E, Ommer B. A variational u-net for conditional appearance and shape generation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8857-8866. 1

[3] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. arXiv preprint arXiv:1406.2661 1

[4] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014 1

[5] Liu Z, Luo P, Qiu S, et al. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1096-1104. 1

[6] Zhonghua Wu, Guosheng Lin, Qingyi Tao and Jianfei Cai1: M2E-Try On Net: Fashion from Model to Everyone. arXiv preprint arXiv:1811.08599 1