Transfer learning for Person Image with Attribute­Decomposed GAN

Member: puda Zhao hao Wang yinhong Qin

# 1 Introduction

The paper[1] that writen by Yifang Men’s team already finished the transform from a certain clothes to other given human pictures by using Attribute­Decomposed GAN[2] (A­D GAN). Processing an intact human picture is a tough work, since directly encodes the entire image might be a tedious task. However, Yifang Men adopted automatic and unsupervised component attributes generator[1] into the frame of GAN[2].

The goal of A­D GAN model is to synthesize high­-quality person images with user­ controlled human attributes, such as pose, head, upper clothes and pants. The corre­sponding keypoint­based can be automatically extracted via an existing pose estimation method[3] .

For improving the generalization ability of texture encoding, inspired by a style transfer method[4] which directly extracts the image code via a pretrained VGG network, A­D GAN model introduce an architecture of global texture encoding by concatenating the VGG features in corresponding layers to its original encoder.

We use a newly discriminator method[5] in this model. This model adapts two discrim­ inators Dp and Dt, and their specific attributes and responsibilities are explained in the discriminator sections. And this model also introduce a new metric called contextual (CX) score, which is proposed for image transformation[6] and uses the cosine distance between deep features to measure the similarity of two non­aligned images, ignoring the spatial position of the features.

# 2 Hypotheses & Research Questions

In the meantime, Attribute­Decomposed GAN works well on human poses transporting. We assume that it is likely to represent attributes of human poses in other data sets. Therefore, we will strive to solve the following research questions:

1. Does it still feasible for other data­set, such as unfashionable data sets, or not three­ dimensional characters pictures, i.e., anime characters?
2. How well it would works with three dimensional characters pictures?
3. Are there any statistical or information-theoretic intuitions behind these methods and experimental results? If so, what are they?
4. Are there any possible solutions to reduce the training cost of this network?

# 3 Dataset

We conduct experiments on the In­shop Clothes Retrieval Benchmark in the Deep­ Fashion database. The dataset containing various poses and clothes, which is of large scale, diversities and quantities. It also has rich annotations, including 7982 number of clothing items.

In the original experiment design, the author randomly pick 101,996 pairs of images for training and 52,712 pairs for testing. We are going to reduce the scale of training to 1000 pairs for training and 75 pairs, because we make transfer learning from the pre­trained model, which could take advantage of the original large­scale model. We will crop images into 172 x 256 resolution which is {I ∈ R3×172×256}, then we generate Component Transfer by human parsing model Look Into Person. Each image is segmented by 8 categories(i.e., background, hair, face, upper clothes, pants, skirt, arm and leg). As for Pose Transfer, we

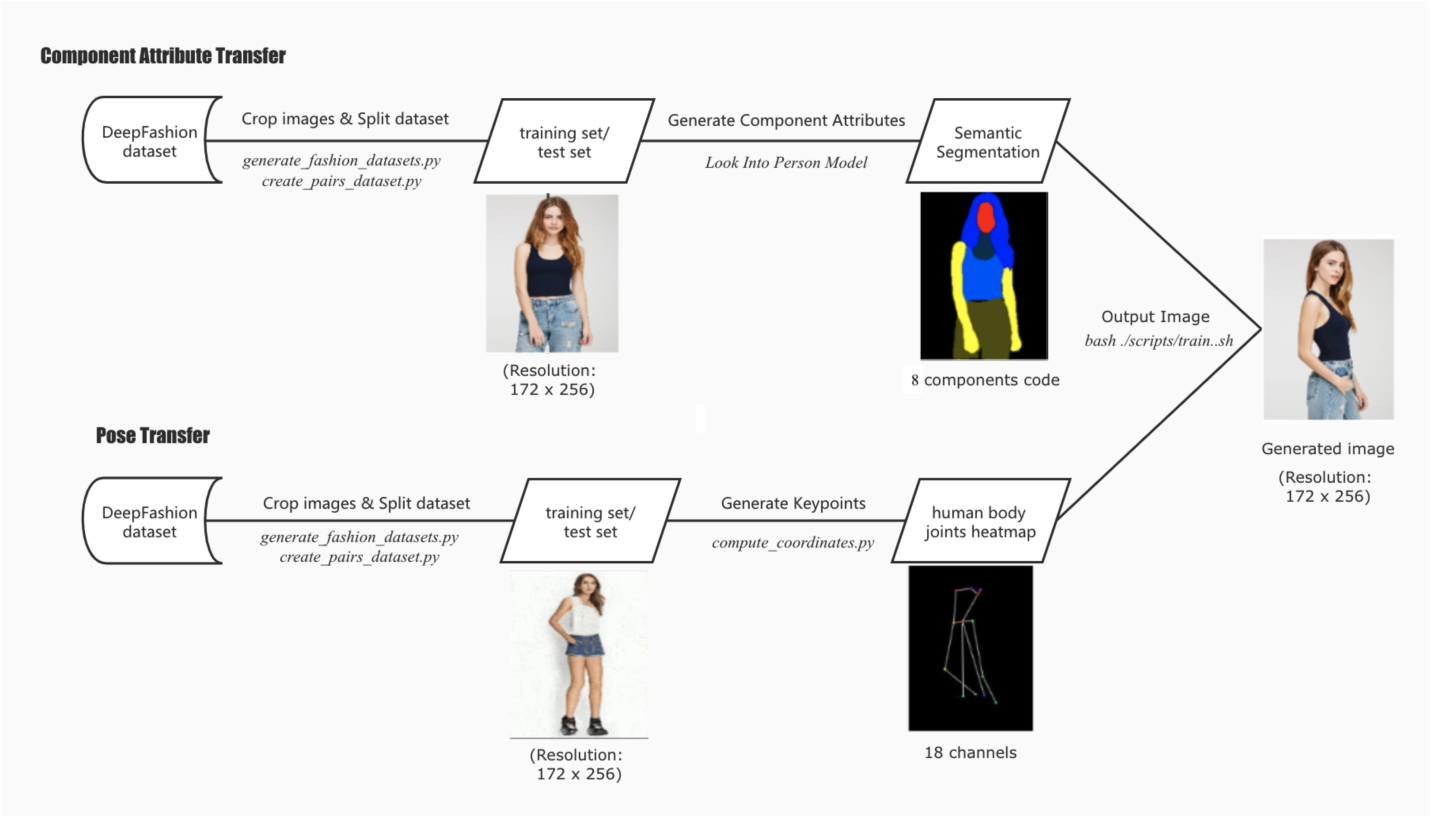
will generate key points of body joints by using OpenPose, the output will be a 18 channel heatmap representing human pose P ∈ R18×172×256 of I. Finally synthesis them together. Starting from the original dataset, the path of data processing is as shown in the following figure 1.

figure 1

And when do the transfer learning, we use 14 3D­anime character pictures as the transfer learning dataset, pair by pair which means 7 × 7 data. Since this model is a really huge model and we only have limited resources, so the dataset could not set too big.

# 4 Deliverables

By the end of this project, we plan to have

1. Achieve component attributes transform model to deepfashion dataset;
2. Transform this model to anime dataset and evaluate the result;
3. Finally create a model which could transform fashion style and pose between different anime characters.
4. Collected all these results in a detailed report and a GitHub repository.

# 5 References

* [1]
* [2]  Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].
* [3]
* [4]  Xun Huang and Serge Belongie. *Arbitrary Style Transfer in Real­time with Adaptive Instance Normalization*. 2017. arXiv: 1703.06868 [cs.CV].
* [5]  Zhen Zhu et al. “Progressive Pose Attention Transfer for Person Image Generation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recog­ nition (CVPR)*. June 2019.
* [6]  Roey Mechrez, Itamar Talmi, and Lihi Zelnik­Manor. *The Contextual Loss for Image Transformation with Non­Aligned Data*. 2018. arXiv: 1803.02077 [cs.CV].