ELE 691 Image and Video Processing
Project Proposal
Motion Transfer based on Generative Adversarial Networks

1 Team

- Kun Wu: kwu102@syr.edu

- Chen Luo: cluo05@syr.edu

2 Task Descriptions

Since the power of the Artificial Neural Networks, deep learning has achieved tremendous success in many tasks about computer vision and image processing. As a seminal generative mode, the generative adversarial network attracts more and more researchers' attention and there are many different methods and tasks that are proposed.

One interesting problem is human motion transfer. Given a training set of target person images with different motions or poses, we aim to train a model that can synthesize novel target person images with arbitrary motions which are different from the training images. The desirable motions are usually from the source person images.

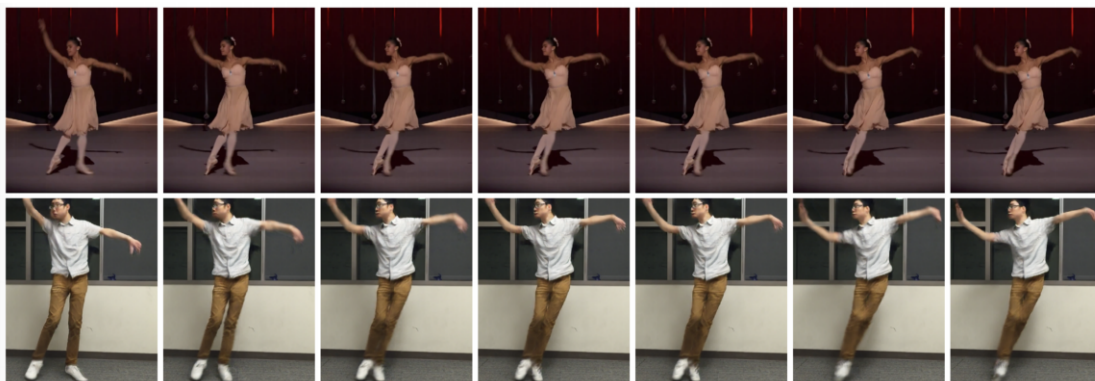An example from (Chan, Ginosar, Zhou, & Efros, 2019) is shown as Figure 1:

Figure 1: **"Do as I Do" motion transfer:** given a YouTube clip of a ballerina (top), and a video of a graduate student performing various motions, our method transfers the ballerina's performance onto the student (bottom). Video: https://youtu.be/mSaIrz8lM1U

Figure 1: The motions of the ballerina are transferred to a target person.

3 Method

In this project, we want to apply the generative adversarial networks on the human motion transfer task. GAN consists of two parts: a generator and a discriminator. The generator takes a kind of input and generates fake images and tries to fool the discriminator and the discriminator tries to tell the fake images

from the real images. Since we wan to transfer different poses to the target person, the generator takes the pose images as input and generates a target person with the corresponding pose. For the pose image, we will use a pre-trained pose estimator Openpose (Cao, Hidalgo, Simon, Wei, & Sheikh, 2018) to extract the keypoints of the source person and draw the pose images by connecting the keypoints with colored lines. An example of the origrinal source person image and pose image are shown Figures 2 and 3:

Figure 2: Source person image

Figure 3: The pose image of the source person

When generating the target person images from the pose images, we want to use the Pix2Pix method (Isola, Zhu, Zhou, & Efros, 2017) which is a well-known method based on GAN.

Contribution: The pix2pix model is a general image-to-image translation model and the paper does not have the experiments about the motion transfer task. And the pix2pix also does not combine the pose estimation technology to do something. So here our main contribution is that we adapt this method on a new task (motion transfer) by combining a pose estimator.

4   Experiments

We will choose the dataset from (Chan et al., 2019) to build our experiments. The dataset contains 5 subjects with thousands of images for each and each image has a 1024*512 resolution. Since there are too many training data and the images are too large, we will choose a part of the 5 subjects and resize the images to train our model. For evaluting our model, we will use MSE and PSNR between the generated images and the real images as the metrics.

References

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). Openpose: realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*.

Chan, C., Ginosar, S., Zhou, T., & Efros, A. A. (2019). Everybody dance now. In *Iccv.*

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Cvpr.*