Motion transfer focuses on creating a plausible image given the source image and arbitrary target pose. Because of the excellent performance shown by the generative adversarial network(ref) and its numerous variations, recent works applied deep learning on motion transfer make great progress. Instead of relying on the multi-camera situation, these deep learning models can be trained on pairs of simple original images and target images and achieve plausible results. (ref https://github.com/budui/Human-Pose-Transfer, +posewarp 5 paper).

(ref Pose Guided Person Image Generation) construct a two-stage framework, which first combines the condition image and target pose into a coarse image of the person and then refines the blurry result in an adversarial way. (ref synthesizing images of humans in unseen poses) segments the complex task into small subtasks. It first does source image segmentation, dividing the original input’s body into different part layers. Then it calculates the spatial transformation between source pose and target pose. Finally, a foreground and background synthesis applied parallelly to generate the final output. (ref Pose-normalized image generation for person re-identification) propose a solution based on generating pose-normalized images for the scalability and generalizability issues. (ref Progressive Pose Attention Transfer for Person Image Generation) utilizes the attention mechanism and progressive style pipeline to address the challenging task of pose transfer. Very recently, (ref Everybody Dance Now) presents a simple video-based method for motion transfer between different people. It trains the model on a specific person with an encoder-decoder-like framework. The well-trained model can then generate videos of this specific person through poses extracted from other people’s motion videos. All these works mentioned above are based on the real human pose transfer, where the condition image is the same material with the output image, while in this work, we try to do a motion transfer with arbitrary anime character and real human pairs.

To extract pose information from input and output image, state-of-the-art pose estimation techniques are widely used in motion transfer. For 2D pose estimation, (ref CPM) first introduces the keypoint heatmap as the representation medium for pose information. (ref hourglass) proposes the idea of learning keypoint heatmap with different scale receptive fields.

The latest work (ref Global Context for Convolutional Pose Machines.) improves original convolutional pose machine architecture and achieves high accuracy with less training data. Because of the friendly interface and well-pretrained model, OpenPose(ref) is plugged in many motion transfer works for real human pose estimation. In this work, we directly use OpenPose interface for real human pose estimation and pretrain (ref Global Context for Convolutional Pose Machines.) for anime pose estimation.

Anime production involves a laborious process of drawing keyframes and anime generative tasks aim to relieve the burden of anime practitioner or offer an easy approach for amateurs to create anime. As a subset of artwork generation, previous works mainly focused on anime face generation (ref Illustration GAN, Anime GAN and Towards the Automatic Anime Characters Creation with Generative Adversarial Networks) and style transfer (ref Style Transfer for Anime Sketches with Enhanced Residual U-net and Auxiliary Classifier GAN, Two-Stage Sketch Colorization). They either generate part of the anime or depend on a strong

Prerequisite such as finely-drawn sketch. With the development of variants of GAN(ref cycleGAN, ACGAN, Progressive GAN, style GAN), (ref full-body anime generation) generates full-body anime character with arbitrary pose gave pose keypoints. However, their model is still trained for a specific anime character. In this work, the anime character is inputted to the model as a condition and thus can generate various anime characters with one-time training.