Consider the traditional pipeline of 2D anime production, it contains laborious work of drawing keyframes of the anime according to the ‘layout’, a sketch indicating what the anime character is doing in every frame. And for amateur anime lovers who want to produce their own anime shorts, it takes great effort to master extra skills and techniques even if they have designed their own anime character in a painting. Fortunately, fast developing generative models like GAN(ref), VAE(ref) make it possible to generate plausible images and videos automatically from the given condition. Besides, the heated synthesis and pose transfer works recent year provide lots of plug-in modules. These all contributes to generate anime shorts without any knowledge of anime production, which give birth to this work.

In this work, we propose an approach to generate an anime video sequence automatically given a real human video sequence and an image of anime character. The real human video performs the motion that we wish to transfer to target anime character image, namely making target anime character do what the real human do in the video. We separately train different phases of transferring and combined them together as a video-to-video translation in testing, enabling practitioners to use it as reference and amateurs to make anime on their own easily.

An image-to-image mapping should be discovered by the whole framework in order to transfer motion from real human to anime video frame by frame. Under this consideration, pairs of real human and anime frame which performs the same motion should be essential if we want to apply supervised learning. However, unfortunately, there are no ready-made datasets for learning this corresponding translation. What’s more, we even cannot find available ready-made dataset for well-placed full-body anime character. The widely-used anime datasets is either for anime face(ref) or messy anime artworks with extremely complex perspectives and composition, which makes the situation even worse. In order to push the framework focusing on the motion transfer, instead of image understanding, and prevent the training process from crashing because of the dirty data, we construct our own real-to-anime motion transfer dataset via 3D anime software MikuMikuDance. Given a real human video and an anime character, we construct a pipeline to automatically generate the corresponding anime video.

It has proved that multi-stage training achieves better performance and also interpretability than one-step training(ref). Keypoint-based pose, which can record the motion signatures and abstract the identity of original input, is widely accepted as intermediate medium for image-to-image motion transfer(ref). The pose estimation for real human is quite mature and there are many off-the-shelf pose detectors such as OpenPose(ref). Therefore, we directly use them to extract pose figures from input real human video. To gap the difference of body size and body ratio between the real human in video and anime character in image, we also need anime character’s pose figure for reference. However, pose estimation for real human and anime characters are not interlinked, which means we cannot directly apply off-the-shelf pose detectors on anime character. We only find one work(ref) special for anime pose estimation but not open source and hence pretrain the state-of-the-art pose estimation work(ref) for anime character using augmented anime pose dataset(ref).

The central contribution of our work is threefold: 1) We construct a new dataset that pairs full-body real human and anime character performing the same motion. 2) We propose an approach about how to get an end-to-end model that transfers motion from real human video to the anime character in the 2D dimension. 3) We dig out suitable state-of-the-art pose estimation and pose transfer work and retrain them using our anime dataset.