

Contents

1 Disentanglement of Shape and Appearance	2
1.1 Disentangling in a Temporal Sequence	2
1.2 Part-wise Disentanglement	2
1.2.1 Part-wise Appearance Transfer	2
1.2.2 Part-wise Shape Changes	2
2 Disentangled Representation	4
3 Bibliography	7

1 Disentanglement of Shape and Appearance

1.1 Disentangling in a Temporal Sequence

Conditional image generation can also be extended to the task of **video-to-video translation**. The two conditioning images can be frames from different videos. One frame is acting as the appearance conditioning and the other as shape conditioning. By generating each frame conditioned on the shape and appearance from two videos, one effectively transfers the appearance of one video to the shape of the other on a frame-to-frame level. We evaluate this frame-to-frame video translation on the BBCPose dataset. The datasets videos of sign language present a delicate and complex articulation of arms and hands. We condition on appearance from videos in the training set and on shape from videos in the test set. A sample for generated frames is shown in Fig. ??, for the complete videos please refer to the supplementary. We want to point out two features of the model here: Firstly, despite no use of smoothing or interpolation between frames the generated sequence is smooth in the temporal domain. This is enabled by a temporally consistent part assignment which is stable across articulation. Secondly, the training on the natural spatial transforms in video data enables the model to encapsulate realistic transitions such as out-of-plane rotation and complex 3D articulation of e.g. hands and even fingers (note the correct translation of the thumbs position in Fig. ??).

1.2 Part-wise Disentanglement

1.2.1 Part-wise Appearance Transfer

The local modelling of parts allows for a part-wise transfer of appearance or shape. In Fig. 2.1 we show the image generation conditioned on a target shape and appearance from a single image, but for several parts the appearance is transferred from another image.

1.2.2 Part-wise Shape Changes

Own Dataset: Move KP, Fig. 1.2

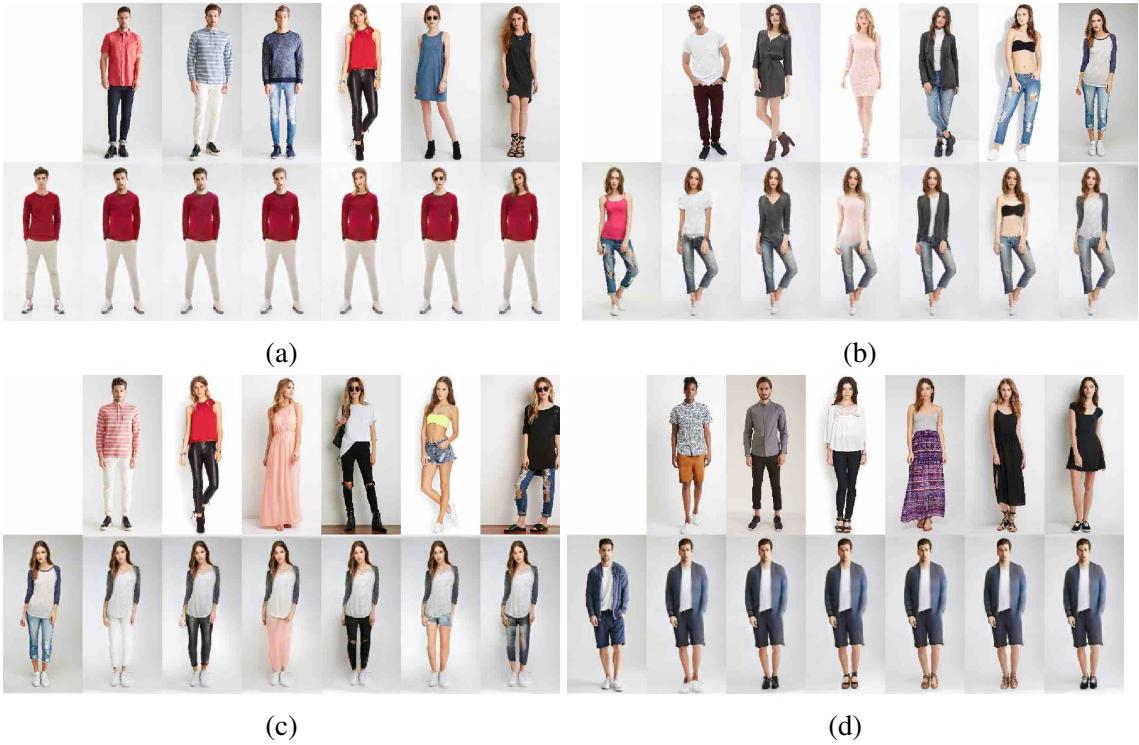


Figure 1.1: Swapping part appearance on Deep Fashion. Appearances can be exchanged for parts individually and without altering shape. We show part-wise swaps for (a) head (b) torso (c) legs, (d) shoes. All images are from the test set.

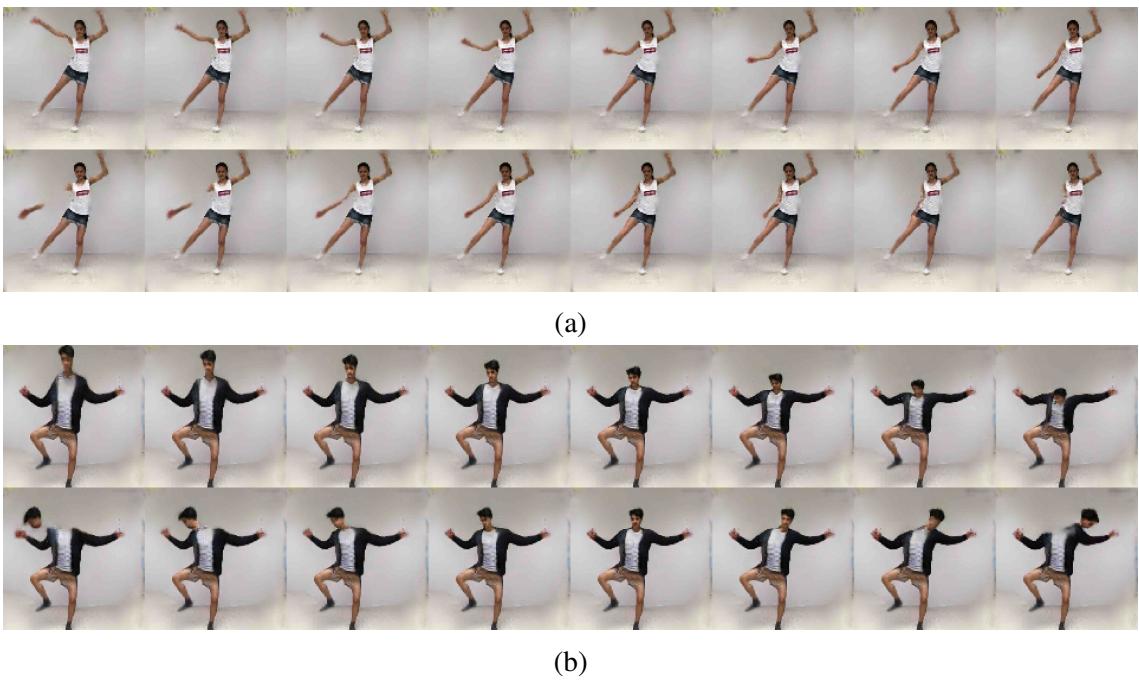


Figure 1.2: Moving individual body landmarks for conditional generation (a) arm (b) head.

2 Disentangled Representation

Local Appearance Transfer. In Fig. 2.1, we show results for successively swapping part appearance on the Deep Fashion dataset.

Video-to-Video Translation. In Fig. 2.2 we show sequences of a frame-to-frame appearance-shape transfer on the BBC Pose dataset. Note that the out-of-plane rotations and fine-grained details of hands and facial expressions are accurately captured. Notice the quality (smoothness, consistency) of the transfer.

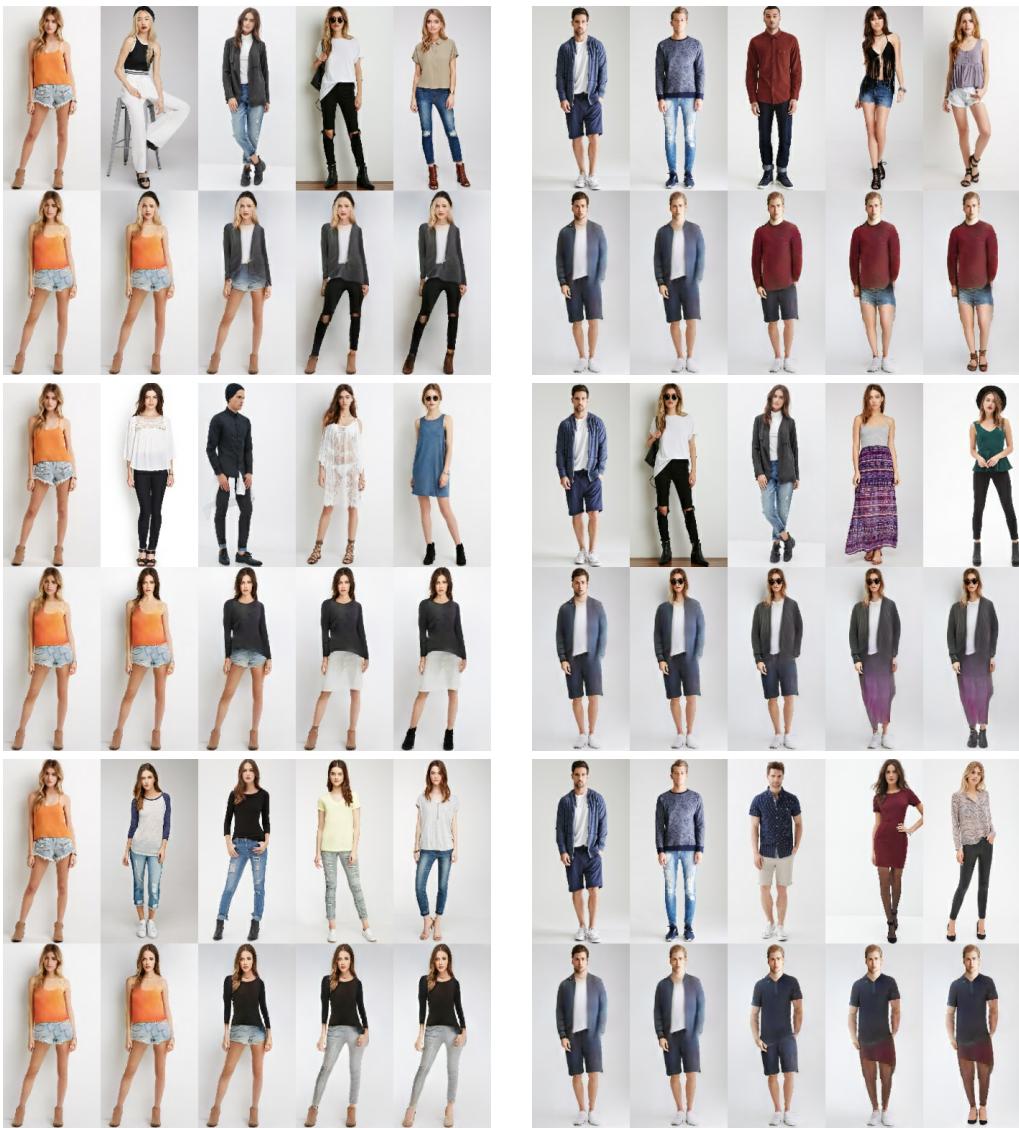


Figure 2.1: Successively altering the appearance of individual parts. We show 6 examples of successively altering appearances of parts using different source images. In each example we start from the original appearance (left-most column). The top row shows ground-truth images (taken from the test-set), which act as the source for the part appearance to be altered. The bottom row then illustrates the new synthesized image, which is generated based on the already altered part appearances plus the current appearance modification. Part appearances are altered in fixed order: head, upper body, legs, feet.



Figure 2.2: Generated sequence on BBC Pose from a target pose sequence (leftmost column) and target appearances (top row).

3 Bibliography