Learning Human Pose from Unaligned Data through Image Translation

1. OVERVIEW

“Learn human-body landmark detectors from unlabeled videos and unaligned annotations”

CONTRIBUTIONS
- Learn landmark detectors from unlabeled videos and unaligned pose annotations. Using no paired data / labelled images.
- Prevent appearance leakage in CycleGAN through (a) novel bottleneck with a differentiable sketch renderer.
- Conditioning the generator on an appearance image.
- Outperform state-of-the-art supervised and unsupervised landmark detectors for human pose.

2. METHOD

ADVERSARIAL ALIGNMENT OF DISCOVERED KEYPOINTS
1. Distill object pose (represented as keypoints) from the input appearance image.
2. Align the discovered object landmarks with human annotations using an adversarial loss.

CycleGAN Appearance Leakage

CycleGAN [Zhu et al., 2017] cheats by encoding appearance information in the geometric representation (skeleton bottleneck).

Geometry Bottleneck

3. RESULTS

Simplified Human3.6M Dataset

Human3.6M Dataset

Pennaction Dataset

4. ABLATIONS

5. DISENTANGLING STYLE & GEOMETRY

Mixing appearance and geometry by conditioning on a different identity.