

Unsupervised Methods for Detecting Object Landmarks

Sangho Lee

MLVU

sangho.lee@snu.ac.kr



Example of unsupervised estimation for human pose

1. Problem to solve

In these days, computer vision seeks to understand object structures that reflect the physical states and representation of objects. And images and videos are an endless source of data, but many of them is hard to be used because of lack of annotations to train. Manual annotations or designs of object structures (e.g., skeleton, semantic parts) are costly and barely available for most object categories, making the automatic representation learning of object structure an attractive solution to this challenge. So, I propose this subject for this project to solve such challenges with unsupervised methods.

2. Related Research

There are several research have tried to solve such problems of lack of annotation with unsupervised methods.

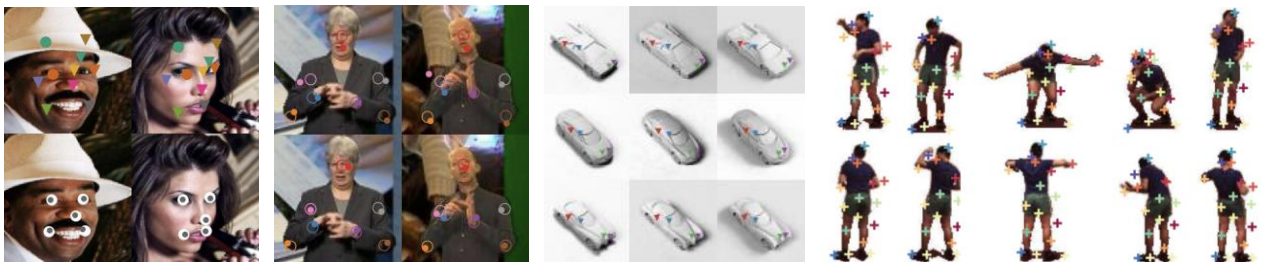
Unsupervised learning of object landmarks by factorized spatial embeddings (ICCV, 2017) proposed an unsupervised method to locate landmarks at the place where a convolutional neural network detects stable visual patterns, but this method does not point landmark at critical location.

Unsupervised Discovery of Object Landmarks as Structural Representations (CVPR, 2018) discovers landmark with a differentiable autoencoder framework for informative landmark detection.

Unsupervised Human Pose Estimation through Transforming Shape Templates (CVPR, 2021) proposed a method for the unsupervised estimation of 2D keypoints requiring only a simple template and an unannotated video of a single human performing actions

Unsupervised Learning of Object Landmarks through Conditional Image Generation (NIPS, 2018) proposed a method for learning landmark detectors for visual objects without supervision

Unsupervised Part-Based Disentangling of Object Shape and Appearance (CVPR, 2019) proposed a method for disentangling object shape and appearance



Example of unsupervised detected landmarks from previous research

3. Dataset and Evaluation

There are plenty of annotated dataset to evaluate the performance of our research and most previous research have evaluated their models by the ratio of detected landmark numbers per whole landmarks.

1) Human pose dataset

- a. DeepFashion: High resolution in-shop clothes images
- b. Penn Action: Human activity videos of sports categories
- c. Human3.6M: Human activity videos

2) Face dataset

- a. CelebA : Human faces

3) Other objects

- a. Cat Head: Nearly 9k images of cat heads
- b. CUB-200-2011: 12k images of birds in the wild

4. Expected Contribution

- 1) Adaptation to other human pose and face related tasks which are lack of labeled dataset such as human pose transfer or fake face generation
- 2) Expand categories to non-human categories from animals to objects like car

5. My Background

Majored in computer science and mathematics from university. Interested in finding representation of images and generating videos in unsupervised ways especially for making source image to follow the action or expression of target video.