Human Pose Estimation: PoseNet

Flavio Amurrio-Moya, G00593001

famurrio@gmu.edu

1   Source Paper

The work done in this paper is based on the 2018 **PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model** [5] written by George Papandreou and Tyler Zhu and Liang-Chieh Chen and Spyros Gidaris and Jonathan Tompson and Kevin Murphy who are part of Gooogle Inc.

The source paper [5] talks about PersonLab which is able to do pose estimation and instance segmentation of people within a picture. This was done with Convolutional Neural Network to detect ceratin keypoints and then use an alogrithm to group these together to identify a person. This model was trained COCO Dataset [4] and it was able to acchive an average precision of 0.665 and 0.687 for single and multiscale infereance [5].

2   Source Code

The work for this paper was done in Javascript with the help of P5.js [1] which a library to easily create interactive projects on the web with most of the capabilities from the browser such as being able to load assets from the web as well as being able to use the webcam for real time capturing. I am also using Ml5.js [2] which is a library that wraps around tensorflow.js [3], both of which allows us to train and model and use it to predict and classify directly on a browser (or any runtime that is able to run javascript). These tools also make use to the graphics card since it is able to compile to something that WEBGL can use and WEBGL can make use of the browser's GPU acceleration. These library also contain the PoseNet model ready to use directly from Gooogle's researchers. The model was opened source but the actual implementation for the training of this model is propiatery making it difficult to replicate the findings. Instead, I will be using the pretrain model directly from Gooogle.

The main reason as to why I picked these tools is mainly for the ease of use and minimal setup as all a person has to do is just add the script tag to their html file.

3 Methodology

The algorithm that PersonLab uses comes in to major steps. First step is to detect all bodies keypoints within an image regardless of where they might be located. This is done thru detecting heatmaps of possible keypoint as well as short-range offset. Then these two do Hough voting to create a Hough array which then gets use with the mid-range offsets to detect the human poses. With the detected human posed and person segmentation mask and long-range offsets, an algorithm is used to infer the instance segmentation.

4 Experiments

4.1 Experiments Description

For this experiment I have first started with the PoseNet model that is built in Ml5.js and used it to give me the the positon of the keypoints. I then drew this keypoints on top of the image to make sure the model was working correctly.

4.2 Source Paper Results

- Clearly identify the results in the source paper that you have replicated or tried to replicate in case you had to make adjustments

4.3 Dataset Used

The dataset used was the COCO dataset from `https://cocodataset.org/#download`. This dataset contained images with the annotated human pose skeleton. Some images contained multiple people. So the features are the image's red, green, and blue values and labels would be the location of the keypoints as well as which keypoints belong together.

4.4 Training and Testing

Since I was using an the google published model, because their approach of how they trained the model was not released, there was no training done.

4.5 Experiment Results

- Tables and/or Plots of the results

4.6 Analysis

References

[1] home | p5.js.

[2] ml5js·Friendly Machine Learning For The Web.

[3] TensorFlow.js | Machine Learning for JavaScript Developers.

[4] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.

[5] George Papandreou, Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, and Kevin Murphy. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model, 2018.