What A Pose: Building a 2D Pose Estimation Model

Archit Jaiswal [1]
[1]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL

This project aims to build a 2D pose estimation model using computer vision techniques. The goal is to develop a system that can accurately detect and track human body poses from 2D images. The project will begin by collecting a dataset of labeled images that depict various human poses. This dataset will be used to train a deep learning model using state-of-the-art techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The model will be designed to predict the location of joints and body parts in the image. To improve the accuracy of the model, additional techniques such as data augmentation and transfer learning will be used. The model's performance will be evaluated using various metrics such as accuracy, precision, recall, and F1-score. The developed model can be used in various applications, such as sports analysis, rehabilitation, and human-computer interaction. For example, the model can be used to track the gesture of a human, allowing the autonomous robot to perform a task corresponding to the operator's gesture. Similarly, the model can be used in rehabilitation to monitor patient progress and ensure that they are performing exercises correctly. Overall, the project aims to contribute to the development of a robust and accurate 2D pose estimation model that can be used in various domains.

I. Project Plan

FINAL PROJECT timeline is of 2 months. On a high level, the project is divided into six stages. The first task is to perform a literature review on pose estimation techniques which will take approximately one week to complete. After the literature review, the second task is to collect the data and label an image dataset. It requires a diverse set of images depicting various human poses and labeling them. Based on the current superficial literature review, COCO (common objects on context) 2017 dataset seems to be a good option to use since it is an open-source dataset [1]. This task is estimated to complete within a week after the completion of a thorough literature review.

Depending on the dataset, the following task will be to perform pre-processing and augmentation before using that data to train the model. Assuming this task will be completed within a week after the dataset is finalized. Once the training data is ready, the plan is to start building a 2D pose estimation model. This task covers the training of a deep learning model using state-of-the-art techniques such as CNNs and RNNs. It also includes experimentation with different architectures and tuning hyperparameters to achieve optimal performance [2]. Accomplishing this complicated milestone will approximately require three weeks. The next task will be to evaluate the model performance using various metrics such as accuracy, precision, recall, and F1-score [2]. This task will cover an analysis of the results and performing necessary adjustments to the model. The deadline to complete this task is the 7th week after the beginning of the project. The final step is to prepare the project report and presentation. The time estimate for this task is about a week. All the milestones described above are scheduled to complete within eight weeks from the beginning of March.

II. Dataset

As briefly described in the previous section, COCO is a large diverse, and reliable dataset that offers multiple crucial features such as object segmentation, recognition in context, and superpixel stuff segmentation covering 250,000 people with keypoints [4]. The COCO dataset contains annotations which can be used to train machine learning models to recognize, label, and describe objects [5].

References

[1] M. Zanetti et al, "Object Pose Detection to Enable 3D Interaction from 2D Equirectangular Images in Mixed Reality Educational Settings," Applied Sciences, vol. 12, (11), pp. 5309, 2022.
[2] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications,* vol. 40, no. 3, pp. 147–156, Mar. 1993.
[3] https://cocodataset.org/home
[4] https://datagen.tech/guides/image-datasets/ms-coco-dataset-using-it-in-your-computer-vision-projects/: :text=What