

College of Engineering Chengannur  
Department of Computer Engineering  
03CS6902 Mini Project  
Abstract of Project Proposed  
EfficientPose: Scalable single-person pose estimation

CHN20MT007, CHN20CSIP03, PRAVEENA K M, praveenamurali1004@gmail.com

May 24, 2021

**Keywords:** 2D human pose estimation, image enhancement, image representation, object detection

**Abstract**

Pose estimation is a computer vision task that infers the pose of a person or object in an image or video. It is a technique used to estimate how a person is physically positioned, such as standing, sitting, or lying down. s.” Realtime multi-person 2D pose estimation is a key component in enabling machines to have an understanding of people in images and videos. This is typically done by identifying, locating, and tracking a number of keypoints or to find the 18 “joints of the body” on a given object or person. For objects, this could be corners or other significant features. And for humans, these keypoints represent major joints like an elbow or knee. There’s also a key distinction to be made between 2D and 3D pose estimation. 2D pose estimation simply estimates the location of keypoints in 2D space relative to an image or video frame. The model estimates an X and Y coordinate for each keypoint. 3D pose estimation works to transform an object in a 2D image into a 3D object by adding a z-dimension to the prediction. 3D pose estimation allows us to predict the actual spatial positioning of a depicted person or object. 3D pose estimation is a more challenging problem for machine learners, given the complexity required in creating datasets and algorithms that take into account a variety of factors – such as an image’s or video’s background scene, lighting conditions, and more. There are two approaches: a bottom-up approach, and a top-down approach. With a bottom-up approach, the model detects every instance of a particular keypoint (e.g. all left hands) in a given image and then attempts to assemble groups of keypoints into skeletons for distinct objects. A top-down approach is the inverse – the network first uses an object detector to draw a box around each instance of an object, and then estimates the keypoints within each cropped region. The OpenPose network [1]

has been one of the most applied HPE methods in real-world applications. It is also the first open-source real-time system for HPE. OpenPose was originally developed for multi-person HPE, but has in recent years been frequently applied to various single-person applications within clinical research and sport sciences. With pose estimation, we're able to track an object or person in real-world space at an incredibly granular level. This powerful capability opens up a wide range of possible applications. In addition to tracking human movement and activity, pose estimation opens up applications in a range of areas, such as: Augmented reality, Animation, Gaming, Robotics. In paper [2] propose a novel convolutional neural network architecture, called EfficientPose, which exploits recently proposed EfficientNets in order to deliver efficient and scalable single-person pose estimation.

## References

- [1] Zhe Cao; Gines Hidalgo; Tomas Simon; Shih-En Wei; Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 128–137, july 2019.
- [2] Heri Ramampiaro Espen AF Ihlen Daniel Groos. Efficientpose: Scalable single-person pose estimation. *Springer journal of Applied Intelligence*, 51:2518–2533, 6, November 2020.
- [3] Gehler P Schiele B Andriluka M, Pishchulin L. 2d human pose estimation: new benchmark and state of the art analysis. In *IEEE Conference on computer vision and pattern recognition (CVPR)*, 2014.

---

Decision: