

Perform Pose Estimation using Computer Vision

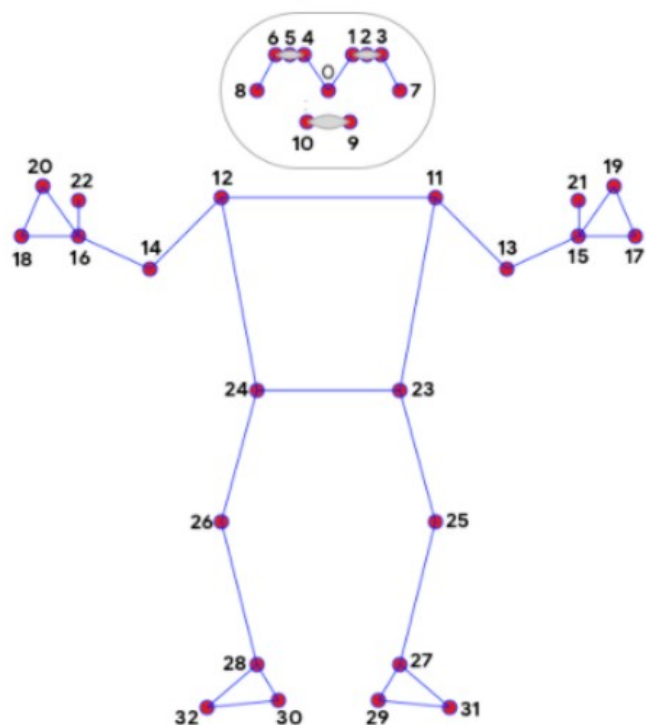
TASK-10

AIRSS1237

Human pose estimation and tracking is a computer vision task that includes detecting, associating and tracking semantic key points. Examples of semantic key points are "right shoulders", "left knees" or the "left brake lights of vehicles".

The performance of semantic key point tracking in live video footage requires high computational resources what has been limiting the accuracy of pose estimation. With the latest advances, new applications with real-time requirements become possible, such as self-driving cars and last-mile delivery robots.

Today, the most powerful image processing models are based on convolutional neural networks (CNNs). Hence, state-of-the-art methods are typically based on designing the CNN architecture tailored particularly for human pose inference .

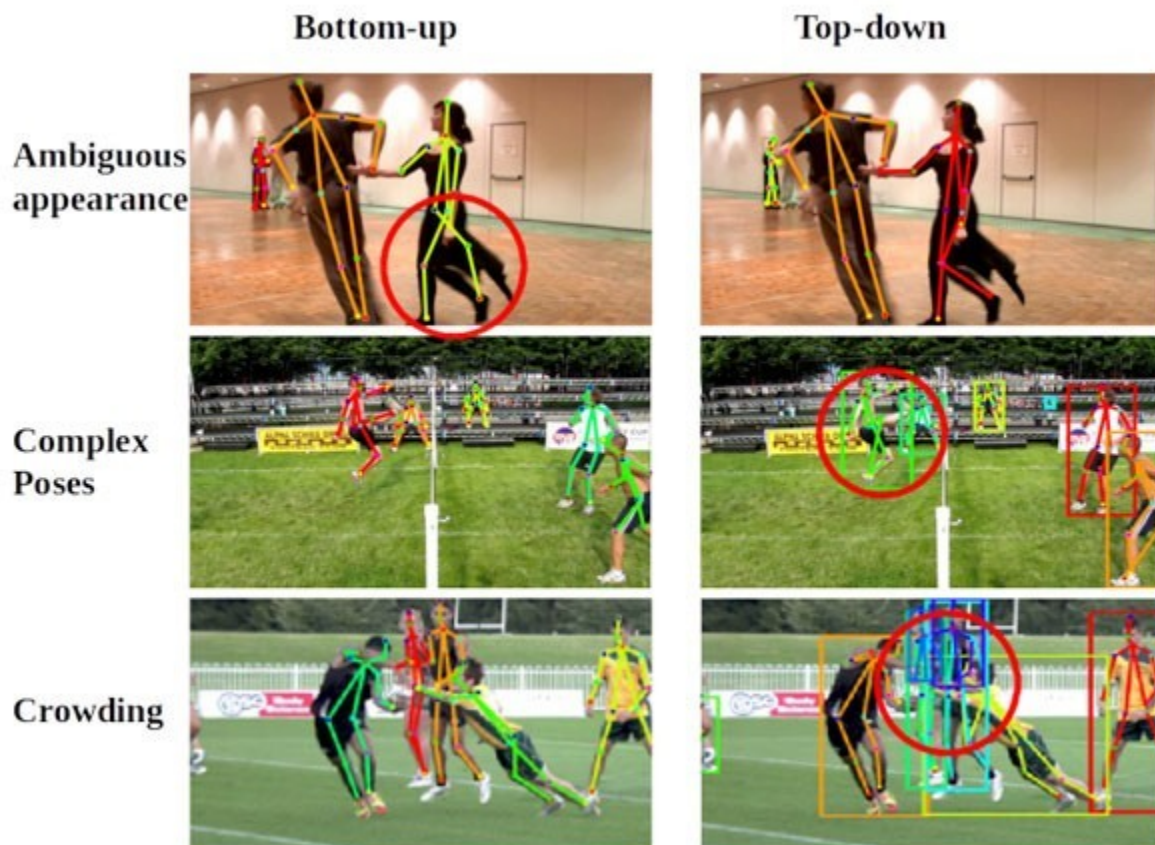


- | | |
|--------------------|----------------------|
| 0. nose | 17. left_pinky |
| 1. left_eye_inner | 18. right_pinky |
| 2. left_eye | 19. left_index |
| 3. left_eye_outer | 20. right_index |
| 4. right_eye_inner | 21. left_thumb |
| 5. right_eye | 22. right_thumb |
| 6. right_eye_outer | 23. left_hip |
| 7. left_ear | 24. right_hip |
| 8. right_ear | 25. left_knee |
| 9. mouth_left | 26. right_knee |
| 10. mouth_right | 27. left_ankle |
| 11. left_shoulder | 28. right_ankle |
| 12. right_shoulder | 29. left_heel |
| 13. left_elbow | 30. right_heel |
| 14. right_elbow | 31. left_foot_index |
| 15. left_wrist | 32. right_foot_index |
| 16. right_wrist | |

Bottom-up VS Top-down methods :-

All approaches for pose estimation can be grouped into bottom-up and top-down methods.

- Bottom-up methods estimate each body joint first and then group them to form a unique pose. Bottom-up methods were pioneered with *DeepCut*.
- Top-down methods run a person detector first and estimate body joints within the detected bounding boxes.



Why do we need Pose Estimation :-

In traditional *object detection*, people are only perceived as a bounding box (a square). By performing pose detection and pose tracking, computers can develop an understanding of human body language. However, conventional pose tracking methods are neither fast enough nor robust enough to occlusions to be viable.

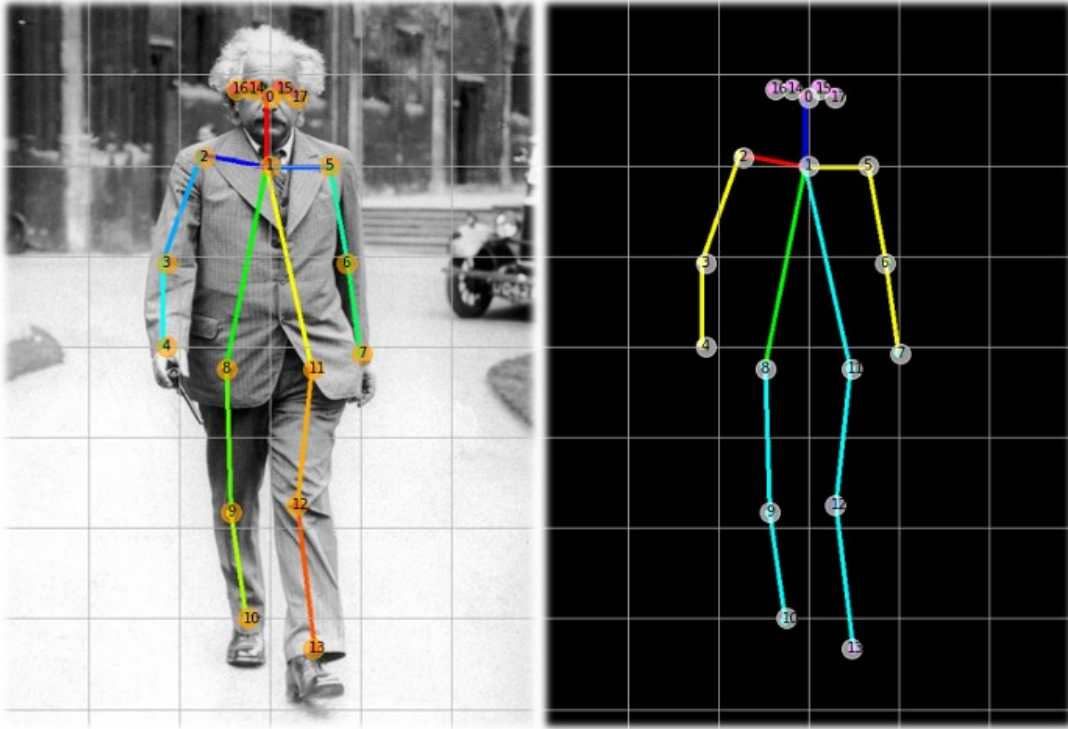
High-performing real-time pose detection and tracking will drive some of the biggest trends in computer vision. For example, tracking the human pose in real-time will enable computers to develop a finer-grained and more natural understanding of human behavior.

This will have a big impact on various fields, for example, in autonomous driving. Today, the majority of self-driving car accidents are caused by "robotics" driving, where the self-driving vehicle conducts an allowed but unexpected stop, and a human driver crashes into the self-driving car. With real-time human pose detection and tracking, the computers are able to understand and predict pedestrian behavior much better – allowing more natural driving.

2D HUMAN POSE ESTIMATION

2D human pose estimation is used to estimate the 2D position or spatial location of human body key points from visuals such as images and videos. Traditional 2D human pose estimation methods use different hand-crafted feature extraction techniques for the individual body parts.

Early computer vision works described the human body as a stick figure to obtain global pose structures. However, modern deep learning based approaches have achieved major breakthroughs by improving the performance significantly for both single-person and multi-person pose estimation. Some popular 2D human pose estimation methods include OpenPose, CPN, AlphaPose, and HRNet.



3D HUMAN POSE ESTIMATION

3D Human Pose Estimation is used to predict the locations of body joints in 3D space. Besides the 3D pose, some methods also recover 3D human mesh from images or videos. It can be applied to various applications, such as 3D animation industries, virtual or augmented reality, and 3D action prediction. 3D human pose estimation can be performed on monocular images or videos (normal camera feeds).

Human Body Modelling

In human pose estimation, the location of human body parts is used to build a human body representation (such as a body skeleton pose) from visual input data. Therefore, human body modeling is an important aspect of human pose estimation. It is used to represent features and key points extracted from visual input data. Typically, a model-based approach is used to describe and infer human body poses and render 2D or 3D poses.

There are three types of models for human body modeling:

- Kinematic Model
- Planar Model

- Volumetric Model

Working of Pose Estimation

Pose estimation utilizes pose and orientation to predict and track the location of a person or object. Accordingly, pose estimation allows programs to estimate spatial positions (“poses”) of a body in an image or video. In general, most pose estimators are 2 steps frameworks that detect human bounding boxes and then estimate the pose within each box.

Pose estimation operates by finding key points of a person or object. Taking a person, for example, the key points would be joints like the elbow, knees, wrists, etc. There are two types of pose estimation: multi-pose and single pose.

- *Single pose estimation* is used to estimate the poses of a single object in a given scene.
- *Multi-pose estimation* is used when detecting poses for multiple objects.

Most popular Pose Estimation Methods

- High-Resolution Net (HRNet)
- Open Pose
- DeepCut
- Regional Multi-Person Pose Estimation (AlphaPose)
- Deep Pose
- PoseNet
- Dense Pose

It's Use applications

1.) Human activity and movement:

Tracking the variations in the pose of a person over a period can also be used for activity, gesture and gait recognition. Some applications could be as below -

- Application to identify if a person follows the exercise regime properly.
- Application to identify the health of a person - in case a person collapses inform somebody close or paramedics.

- Application to identify body language and flag suspicious individuals to nearby authorities - used at airports.

2.) Augmented reality:

CGI Application to track the human pose variations to render graphical animations - eg Thanos.

3.). Robotics:

Robotics can be taught to mimic human poses, activities by tracking and following human instructor demonstrations, instead of manually programming robots.