OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields

OpenPose can provide real-time 2D pose estimation using a nonparametric representation to learn the body parts in an image dataset. This is the first open source library available real time system for multi-person 2D pose detection, including body, foot, hand, and facial key points.

The method takes the entire image as the input for a CNN to jointly predict confidence map for body part detection and PAFs for part association. The parsing step performs a set of bipartite matchings to associate body part candidates. Finally assemble them into full body poses for all people in the image.

Human estimation has largely focused on finding body parts of individuals. Inferring the pose of multiple people in images presents a unique set of challenges.

First, each image may contain an unknown number of people that can appear at any position or scale.

Second, interactions between people induce complex spatial interference, due to contact, occlusion, or limb articulations, making association of parts difficult.

Third, runtime complexity tends to grow with the number of people in the image, making real-time performance a challenge.

In this paper, the authors present an efficient method for militiaperson pose estimation with competitive performance on multiple public benchmarks.

They present the first bottom up representation of association scores via Part Affinity Fields (PAFs), a set of 2D vector fields that encode the location and orientation of limbs over the image domain.

First, a feedforward network predicts a set of 2D confidence maps of body part locations and a set of 2D vector fields of part affinity fields (PAFs), which encode the degree of association between parts.

Finally, the confidence maps and the PAFs are parsed by greedy inference to output the 2D key points for all people in the image.

Given a set of detected body parts, the model needs to assemble them to form the full-body poses of an unknown number of people.

It need to confidence measure of the association for each pair of body part detections, i.e., that they belong to the same person.

The problem of finding the optimal parse corresponds to a K-dimensional matching problem that is known to be NP-Hard. In this paper, they present a greedy relaxation that consistently produces high-quality matches.

With all limb connection candidates, they can assemble the connections that share the same part detection candidates into full-body poses of multiple people. The

optimization scheme over the tree structure is orders of magnitude faster than the optimization over the fully connected graph.

The current model also incorporates redundant PAF connections (e.g., between ears and shoulders, wrists and shoulders, etc.). This redundancy particularly improves the accuracy in crowded images. To handle these redundant connections, they slightly modify the multi-person parsing algorithm. While the original approach started from a root component, they algorithm sorts all pairwise possible connections by their PAF score. If a connection tries to connect 2 body parts which have already been assigned to different people, the algorithm recognizes that this would contradict a PAF connection with a higher confidence, and the current connection is subsequently ignored.

Real-time multi-person 2D pose estimation is a critical component in enabling machines to visually understand and interpret humans and their interactions. In this paper, they present an explicit nonparametric representation of the key point association that encodes both position and orientation of human limbs.

Second, they design an architecture that jointly learns part detection and association.

Third, they demonstrate that a greedy parsing algorithm is sufficient to produce high-quality parses of body poses, and preserves efficiency regardless of the number of people.

Fourth, they prove that PAF refinement is far more important than combined PAF and body part location refinement, leading to a substantial increase in both runtime performance and accuracy.

They have open-sourced this work as OpenPose [4], the first real-time system for body, foot, hand, and facial key point detection.