e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

POSE ESTIMATION TECHNIQUES USING OPENPOSE, POSENET AND HRNET

Ramtenki Santhosh*1, Vasam Ganga*2

*1,2Master Of Technology In Computer Science, Department Of Computer Science And Engineering, JNTUH University College Of Engineering Science And Technology, Hyderabad, Telangana, India.

DOI : https://www.doi.org/10.56726/IRJMETS61308

ABSTRACT

By performing posture discovery and posture following, computers can create an understanding of human body dialect. However, conventional pose tracking methods are neither fast enough nor robust to occlusions. High-performing genuine- time posture discovery and following will drive a few of the greatest patterns in computer vision. This paper presents various key point detection techniques analyzes each one and presents the so far best performed technique that can solve a huge sum of problems in understanding the object perfectly.

Keywords: OpenPose, PoseNet, HRNet.

I. INTRODUCTION

In traditional object discovery, people are only perceived as a bounding box(a forecourt). This will have a big impact on colorful fields, for illustration, independent driving, sports, healthcare, and numerous further. moment, the maturity of tone-driving auto accidents are caused by “robotic” driving, where the tone-driving vehicle conducts an allowed but unanticipated stop, and a mortal motorist crashes into the tone-driving auto. With real- time body pose discovery and shadowing, the computers are suitable to understand and prognosticate rambler geste.

Much better - allowing further natural driving. mortal disguise estimation aims to prognosticate the acts of mortal body corridor and joints in images or vids. Since disguise movements are frequently driven by some specific mortal conduct, knowing the body disguise of a human is critical for action recognition and videotape understanding. Detecting the mortal disguise is a grueling task because the body’s appearance stoutly due to different forms of apparel, arbitrary occlusion, occlusions due to the viewing angle, and background surrounds. Pose estimation requirements to be robust to grueling real- world variations similar as lighting and rainfall. thus, it's challenging for image processing models to identify fine- granulated common equals. It's especially delicate to track small and slightly visible joints. Pose estimation utilizes disguise and exposure to prognosticate and track the position of a person or object. Consequently, the ML fashion allows programs to estimate spatial positions(“acts”) of a body in an image or videotape. In general, most pose estimators are 2-way fabrics that descry mortal bounding boxes and also estimate the disguise within each box. Pose estimation operates by chancing crucial points of a person or object. Taking a person, for illustration, the crucial points would be joints like the elbow, knees, wrists, etc. There are two types multi-pose and single-disguise. Single-disguise estimation is used to estimate the acts of a single object in a given scene, while multi-pose estimation is used when detecting acts for multiple objects.

II. LITERATURE REVIEW

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

Authors present here a real-time approach to detect 2D pose regarding multiple people in an image. The proposed methodology is based on a nonparametric representation, so-called Part Affinity Fields (PAFs), which learn to associate body parts with individuals in the image. This bottom-up system is able to attain high accuracy and real-time performance, whether there is a single person or hundreds of people in the image. That work has been released as OpenPose, the first open source real-time system for multi-person 2D pose detection, including body, foot, hand, and facial key points. By Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, Yaser Sheikh, 2019.

Adversarial PoseNet: A Structure-aware Convolutional Network for Human Pose Estimation.

Here, they introduce a new structure-aware convolutional network that can implicitly consider such priors athe time of training of the deep network. Explicit learning of such constraints is usually difficult. Instead, we design

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

discriminators that are targeted to distinguish between real or fake poses, for example, those that are biologically unrealistic. If the pose Generator G generates results which the discriminator fails to distinguish from real ones, then again the network successfully learns the priors. In order to better capture the structure dependency of human body joints, the generator G is designed in a stacked multi-task manner for the prediction of poses and occlusion heatmaps simultaneously. Then the pose and occlusion heatmaps are fed into the discriminators to predict whether the pose is real. The training strategy of the network follows that of the conditional GANs. (Yu Chen, Chunhua Shen, Xiu-Shen Wei, Lingqiao Liu, Jian Yang 2017).

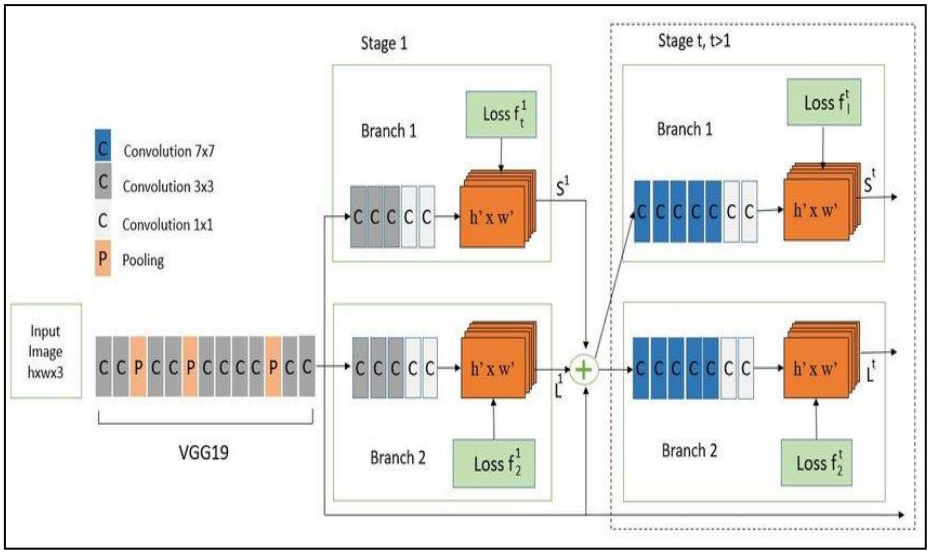
Deep High-Resolution Representation Learning for Human Pose Estimation.

It starts from a high-resolution sub-network as the first stage, gradually adds more and more high-to-low resolution sub-networks one by one to form more stages, and connects the multi-resolution sub-networks in parallel. Atrous spatial pyramid pooling: Increased multi-scale fusions where all the high-to-low resolution representations receive information from other parallel representations many times over, yielding rich high-resolution representations. Thereby, the predicted key-point heatmap is potentially more accurate and spatially more precise. KeSun, Bin Xiao, Dong Liu, Jingdong Wang 2019.

III. METHODOLOGY

OpenPose:

The OpenPose library originally pulls out features from a picture using the first many layers. The extracted features are then inputted into two parallel divisions of convolutional network layers. The first division predicts a set of 18 confidence maps - with each of them denoting a specific part of the human pose skeleton. The coming branch predicts another set of 38 Part Affinity Fields(PAFs) that denotes the position of association between corridor. The after stages are used to clean the prognostications made by the branches. With the help of confidence charts, dual graphs are made between dyads of corridor. PAF values, weaker links are pruned inthe bipartite graphs. Now, applying all the given steps, human pose skeletons can be estimated and allocated to every person in the picture.



Open Pose Architecture

Overview of the Pipeline: The OpenPose Pipeline is made up of several tasks that are completed in order:

a). Acquisition of the entire image as input (image or video frame).

b). Two-branch CNNs jointly predict confidence maps for body part detection.

c). Estimate the Part Affinity Fields (PAF) for parts association.

d). Set of bipartite matchings to associate body parts candidates.

e). Assemble them into full-body poses for all people in the image.

Features: The OpenPose human posture detection library has many features; the most important among them are listed below.

1. Real-time 3D single-person key point detection.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

2. Real-time 2D multi-person key point detections.

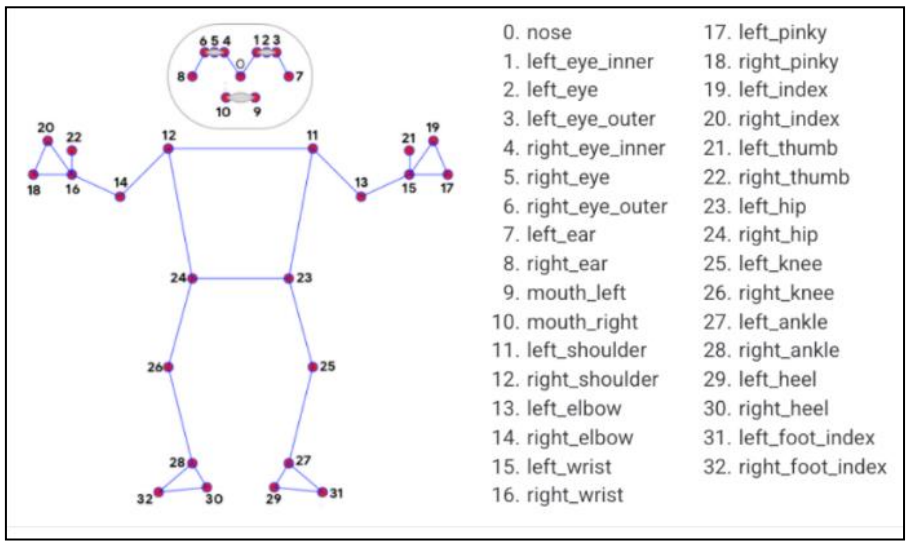
3. Camera parameters, intrinsic, extrinsic, and distorted.

Limitations:

The major problem with OpenPose is the low resolution of the produced outputs limits the amount of information that could be embedded in key point estimates. Therefore, OpenPose is less appropriate for use in applications like elite sports and medical evaluations where great precision in movement kinematics assessment is required. network for HPE performed by a single person using marker-free movements.

PoseNet:

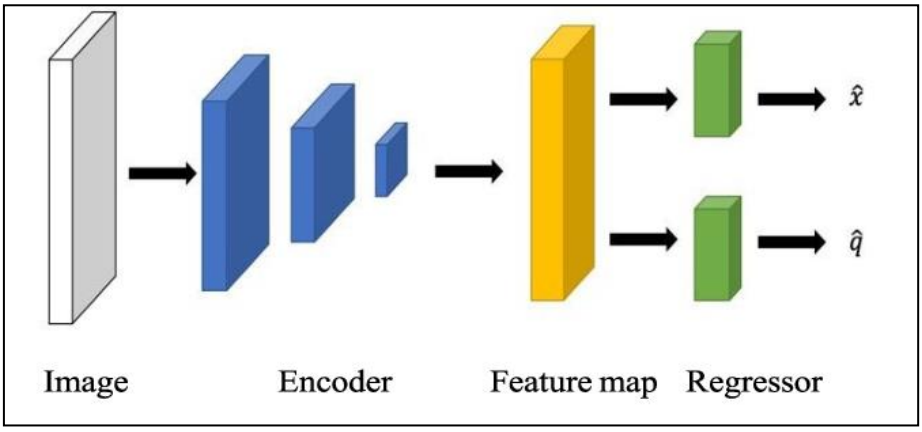
Pose estimation refers to the process for determining the spatial positions of key body joints in the human figure. A deep learning model, by the name PoseNet makes predictions of human pose with the use of CNN architecture. This network is trained on a large data set of annotated images and videos of people in different poses. This will allow PoseNet to learn associations among the body parts and their respective configuration; hence, it can effectively calculate the pose from an input image or any video frame of a person.



Key Points in PoseNet

Architecture and Operation:

The architecture of a deep neural network is used by PoseNet. The strength of PoseNet emanates from deep learning supported by convolutional layers. It takes an image as input and returns a set of key points together with associated confidence scores, conveying information about the exactness of every key point's location. Essentially, it is made up of backbone architecture for feature extraction, followed by several more layers in charge of predicting the key points. More importantly, these key points continue to connect in order to form the skeleton of a human body, and it hence effectively maps the coordinates of the 2D image onto the body part.



PoseNet Architecture

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

Limitations and Improvements:

While computer vision grows day by day, the prospects and applications of PoseNet also grow. A few of the possible themes that might set the trend in the future for it could be as follows:

3D Pose Estimation: This would involve increasing the capability of PoseNet to include the estimation of 3D pose-something quite useful in augmented reality and robotics applications.

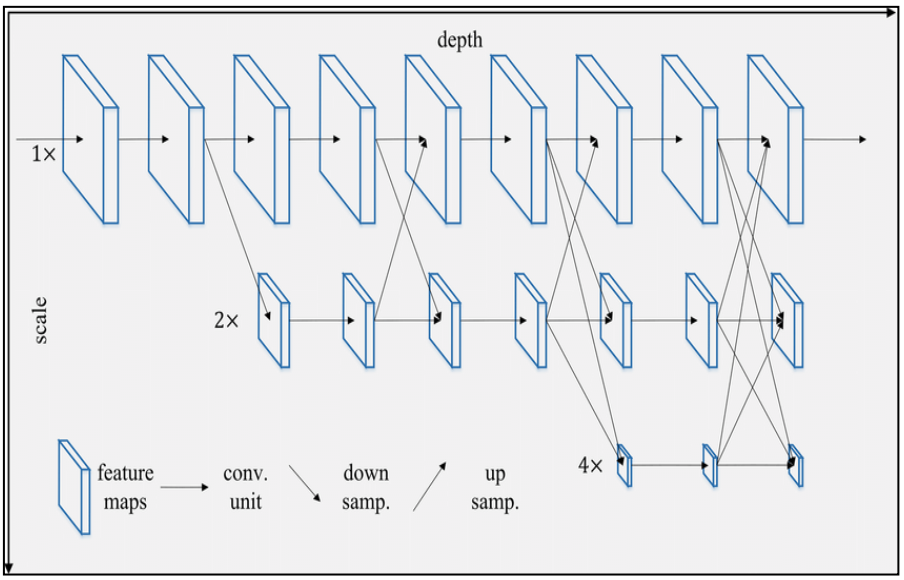
Real-time performance: It needs to be further optimized in terms of speed and efficiency to get real-time pose estimation on resource-constrained devices like smartphones and edge devices.

Custom Pose Detection: Users can train PoseNet for their specific poses, gestures, or body parts, which renders it even more versatile and adaptable.

Multimodal Integration: Merging PoseNet with other sources of sensory data, like depth from LiDAR or radar, for improved accuracy and robustness in challenging environments.

HRNet:

The proposed HRNet maintains high-resolution representations throughout the process. We start from a high-resolution convolution stream and add high-to-low resolution convolution streams one by one. The multiresolution streams are connected in parallel. We obtain a network consisting of several (four in the current design) stages, and the nth stage contains n streams corresponding to n resolutions. We conduct multi-resolution fusions, exchanging the information across the parallel streams over and over.



HRNet Architecture

The semantic strength and the spatial precision of the high-resolution representation learned from HRNet lie in the following two aspects: First, our approach connects the convolution streams of high-to-low resolution in a parallel manner rather than in a serial one. Therefore, compared with recovering the high resolution from the low resolution, our approach can maintain the high resolution directly, and thus the learned representation is spatially more precise. While most of the existing schemes fuse the high-resolution low-level and upsampled low-resolution high-level representations, we propose performing repetitive multi-resolution fusions to enhance higher solution representations using the low-resolution representations and viceversa. Thus, all the representations of high-to-low resolution become semantically stronger.

Features and Improvements:

Here, we present a high-resolution network for human pose estimation that naturally yields accurate and spatially precise key point heat maps. The effectiveness lies in two folds: the network maintains high resolution from the first layer to the last without recovering high resolution and it repeatedly fuses multi-resolution representations, which makes the high-resolution representation reliable. The applications to other dense prediction tasks, such as face alignment, object detection, and semantic segmentation, and investigating how to aggregate multi-resolution representations in a lighter way are the works to be done in the future.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

IV. RESULTS AND DISCUSSION

After successful analysis, the results come out to be as follows:

HRNet:

Key Features: It is known for maintaining high-resolution representations throughout the network. Due to this nature, it could capture very minute details related to body parts.

Advantages: It captures fine-grained details of human pose estimation tasks really well. Therefore, it will be suitable for those applications where high precision is required.

Applications: HRNet finds its applications in several computer vision tasks, such as estimating human pose, recognition of actions, and localization of key points.

OpenPose:

Key Features: OpenPose is an open-source library that detects and tracks multiple key points of the human body, including hand and face.

Advantages: It's versatile to be applied to real-time pose estimation in many scenarios. Also, OpenPose is capable of multi-person pose estimation.

Applications include human-computer interaction, augmented reality, and motion capture. OpenPose is used in applications like human-computer interaction, augmented reality, and motion capture.

PoseNet:

Key Features: PoseNet is another lightweight model from Google. Running in real-time in web browsers is quite possible. It is also part of TensorFlow.js.

Advantages: PoseNet is rather lightweight; hence, it can run without problems on devices with rather limited computational resources. Thus, it's suitable for web-based applications.

Applications: PoseNet is usually employed in applications of browsers, web augmented reality, and interaction that involves real-time pose estimation.

Considerations for Multi-Person Pose Estimation:

Accuracy vs. Speed: HRNet is generally more accurate than the two, since it is high-resolution, whereas PoseNet generally takes real time into consideration. OpenPose balances both.

Computational Resource: The computational needs of both HRNet and OpenPose may be higher than those of lightweight PoseNet, which is intended to run on web applications. Deployment Platform: Since PoseNet can execute in web browsers, their applications can be executed from web-based platforms. The other two might be more accustomed to being deployed on server-based or offline platforms.

Conclusion Eventually, the choice among HRNet, PoseNet, and OpenPose should be made with consideration of specific requirements that an application might have due to the aspects: accuracy, real-time performance, and computational resources.

V. CONCLUSION

Many factors would come into play in order to make a decision.

It has been indicated that HRNet acts as a state-of-the-art model in high-resolution tasks where minor details count.

Real-time performance: PoseNet naturally fits the lightweight requirements, hence it can also be applied in situations where computational resources are more restricted or for real-time performance.

Multitude: OpenPose has been developed to focus on estimating the poses of multiple persons; hence, it is versatile in many ways.

A decision between HRNet, OpenPose, and PoseNet would depend on the exact requirements of your application, the amount of detail you need, and how much computing power you have in the application. Each one of them has strengths and weaknesses; hence, choosing between them needs not only consideration of a particular use case, but also a number of performance factors.

ACKNOWLEDGEMENT

We extend our sincere gratitude to the authors of the research papers reviewed in this study. We are also thankful to our academic mentors for their invaluable guidance and feedback.

VI. REFERENCES

[1] X. Nie, J. Feng, J. Xing, and S. Yan, "Pose partition networks for Multi-person pose estimation," in ECCV, 2018

[2] V. Belagiannis and A. Zisserman. Recurrent human pose estimation. In Proc. IEEE Int. Automatic Face & Gesture Recognition, 2017.

[3] OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields 30 May 2019.

[4] "Adversarial PoseNet: A Structure- aware Convolutional Network for Human Pose Estimation" Yu Chen Chunhua Shen Xiu Shen Wei Lingqiao Liu Jian Yang Nanjing University of Science and Technology University of Adelaide.

[5] Deep High-Resolution Representation Learning for Human Pose Estimation 25 Feb 2019/

[6] W. Yang, W. Ouyang, H. Li, and X. Wang. End-to end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation. In CVPR, pages 3073–3082, 2016.

[7] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In CVPR, pages 1385–1392, 2011.

[8] T. Zhang, G. Qi, B. Xiao, and J. Wang. Interleaved group convolutions. In ICCV, pages 4383–4392, 2017.

[9] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In CVPR, pages 6230–6239, 2017.

[10] L. Zhao, M. Li, D. Meng, X. Li, Z. Zhang, Y. Zhuang, Z. Tu, and J. Wang. Deep convolutional neural networks with merge-and-run mappings. In IJCAI, pages 3170–3176,