Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

A review of human posture recognition

Shugang Liu[1], Wei Guo[2,] *

[1]School of North China Electric Power University, Baoding, China

[2]School of North China Electric Power University, Baoding, China

*Corresponding author Email: wileyya@163.com

Abstract: Human posture recognition is a research hotspot in the field of computer vision and human-computer interaction. This article first briefly introduces the classification of human posture recognition, then focuses on the method of human posture recognition, and gives a detailed introduction to the development of human posture recognition. Finally, It summarized the current problems and looked forward to the future development.
Keywords: Human posture recognition, deep learning, network structure, single pose recognition, multi-person pose recognition.

---

1. INTRODUCTION

In recent years, with the rapid development of computer vision technology, human posture recognition technology has gradually become a research hotspot. The goal of human posture recognition is to detect the key point information of the human body. With the key point information, the position information of the human body in the current image data can be determined. As a milestone technology in the field of human-computer interaction, human posture recognition technology has important application prospects in human motion analysis, medical health analysis, film and television animation games, etc. This article mainly summarizes the development of human posture recognition technology in recent years.

2. ORGANIZATION OF THE TEXT

2.1 Classification of human posture recognition

Human posture recognition can be divided into two categories: single-person pose recognition and multi-person pose recognition.

Single-person pose recognition means that the input image contains only a single human body instance, and only the key point position of a single human body in the input image is detected.

Multi-person pose recognition means that the input image contains multiple human instances, and it is necessary to detect multiple human key points in the image and perform accurate human body division.

2.2 Human posture recognition method

Human posture recognition methods can be divided into methods based on graph structure models and methods based on deep learning. Fig. 1 is the organizational structure of this chapter.

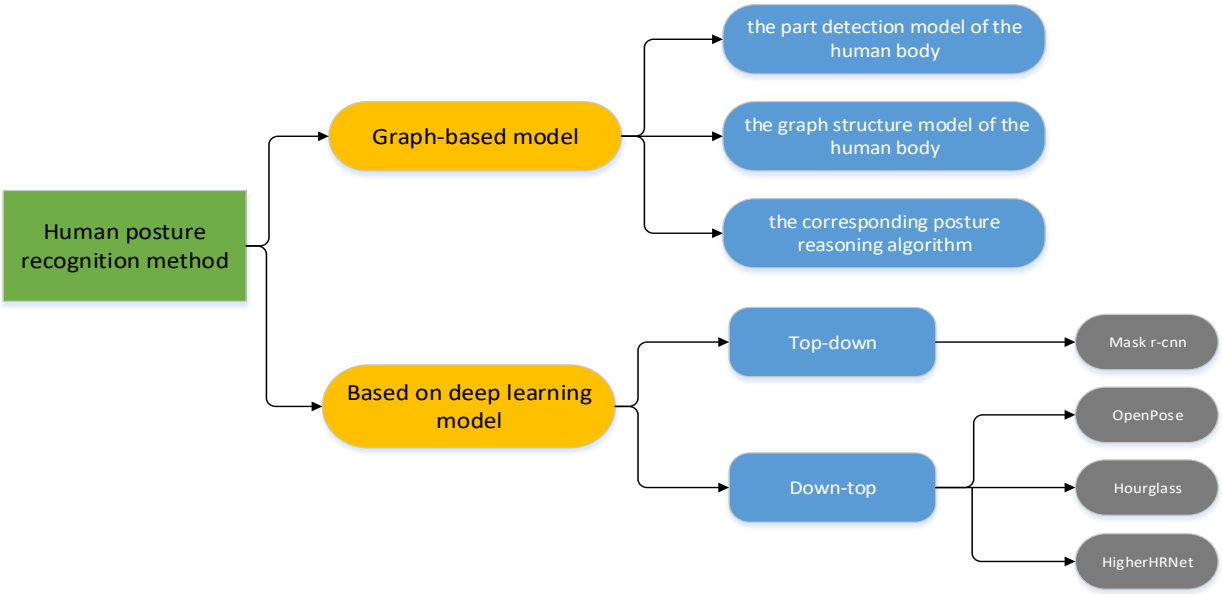

Fig. 1 The organizational structure of this chapter

2.2.1 Human posture recognition method based on graph structure model

The human posture recognition method based on the graph structure model can be divided into three parts: the part detection model of the human body, the graph structure model of the human body and the corresponding posture reasoning algorithm. This method firstly extracts the features of the limb parts of the human body; secondly, it models the relationship between the parts of the human body. However, because the human body has the characteristics of flexible parts and diverse postures, a large number of models need to be pre-built. It makes the model's parameters huge. In addition, the method based on graph structure model is difficult to solve the occlusion problem, which makes the posture search process slower.

2.2.2 Human posture recognition method based on deep learning

Since the rise of deep learning, methods based on deep learning have attracted more attention due to their advantages in omitting the complexity of manual design features and improving the accuracy and efficiency of feature descriptions, and they have also developed rapidly in recent years. Nowadays, the accuracy and efficiency of single-person posture recognition technology has been done very well, and more and more people have begun to study multi-person posture estimation technology.

Multi-person pose recognition technology can be divided into a top-down method and a bottom-up method according to its detection direction.

1. Top-down

The top-down algorithm is currently a network architecture with the highest recognition accuracy in multi-person human pose estimation. The posture recognition process can be divided into two stages. The first stage is to use a human body detector (target recognition algorithm) to detect and crop the human target in the image, and the second stage is to use a single human pose estimation algorithm to recognize the key point information of the human body.

Mask R-CNN [1] is a typical target detection network. A mask branch is added on the basis of Faster

R-CNN [2]. At the same time, replacing Roipool with ROIAlign greatly improves the accuracy of target segmentation detection. Mask R-CNN [1] can be extended to human posture recognition. For single pose recognition, Mask R-CNN [1] performs one-hot encoding on the positions of k key points, using Mask R-CNN [1] for K Each mask predicts its type, recognizes key points accurately, and achieves pixel-level segmentation.

Mask R-CNN [1] can also be broadened to the field of multi-person pose recognition. Detect each generated candidate area. When it is detected that the area contains a human type, the position of each key point on the human body will be one-hot encoded.

The top-down multi-person posture recognition method can improve the accuracy of recognition, but because its computing time is closely dependent on the number of people in the image, as the number of people in the image increases, the time for gesture recognition is gradually Increase.

2. Bottom-up

The bottom-up algorithm is completely different from the top-down algorithm. The principle is to first detect all the key points in an image, and then group the detected key points to find the key points belonging to each person.

The OpenPose[3] algorithm proposed by CVPR2017 is currently one of the most popular bottom-up multi-person human posture estimation algorithms. Its network structure mainly improves CPM[4], and its core lies in the Part Affinity Field (PAF) proposed in the article. The principle is to input an image and pass 7 stages to get PCM and PAF. Then a series of even matches are generated according to PAF. Due to the vector nature of PAF, the generated even matches are very correct, and finally merged into a person's overall skeleton. AE [5] and CenterNet [6] are also representative networks in multi-person posture recognition. Both use Hourglass [7] network as its main part, and use multi-scale representation fusion to obtain more accurate joint point information. Fig. 2 shows an example of a single "hourglass" module. However, the operation of multi-scale representation fusion used by them is only carried out at the end of the network, ignoring the opportunity for mutual fusion of representations in the middle of the network and promotion of improvement. Therefore, these methods are difficult to predict the correct posture for the small human body.

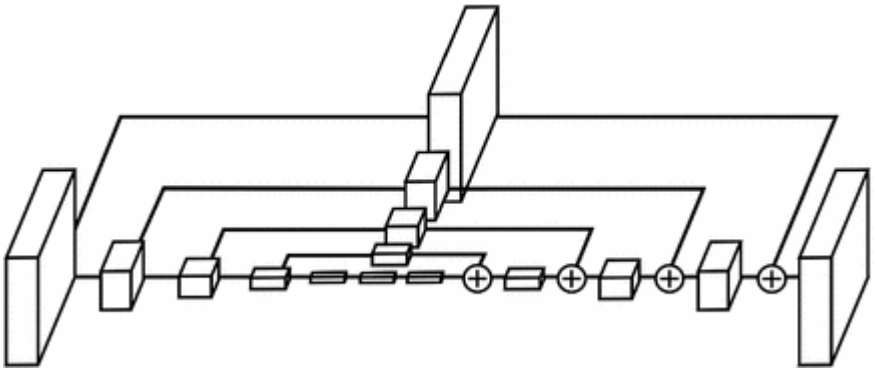

Fig. 2 An illustration of a single “hourglass” module [7]

HigherHRNet [8] proposed by CVPR2020 is a new bottom-up human pose estimation method for learning scale perception representations using high-resolution feature pyramids. This method is equipped with multi-resolution supervision for training and multi-resolution aggregation for inference, which can solve the challenge of scale changes in bottom-up multi-person pose estimation, and can locate key points more accurately, especially For the little ones. The feature pyramid in

HigherHRNet[8] includes the feature map output of HRNet[9] and the high-resolution output that is up-sampled by transposed convolution. In the COCO test-dev, the AP performance of HigherHRNet's medium human body is 2.5% higher than the previous best bottom-up method, showing its effectiveness in dealing with scale changes. In addition, HigherHRNet [8] obtained the latest results on COCO test-dev (AP: 70.5%) without using optimization or other post-processing techniques, thus surpassing all existing bottom-up methods. Fig. 3 is the HigherHRNet network illustration diagram.

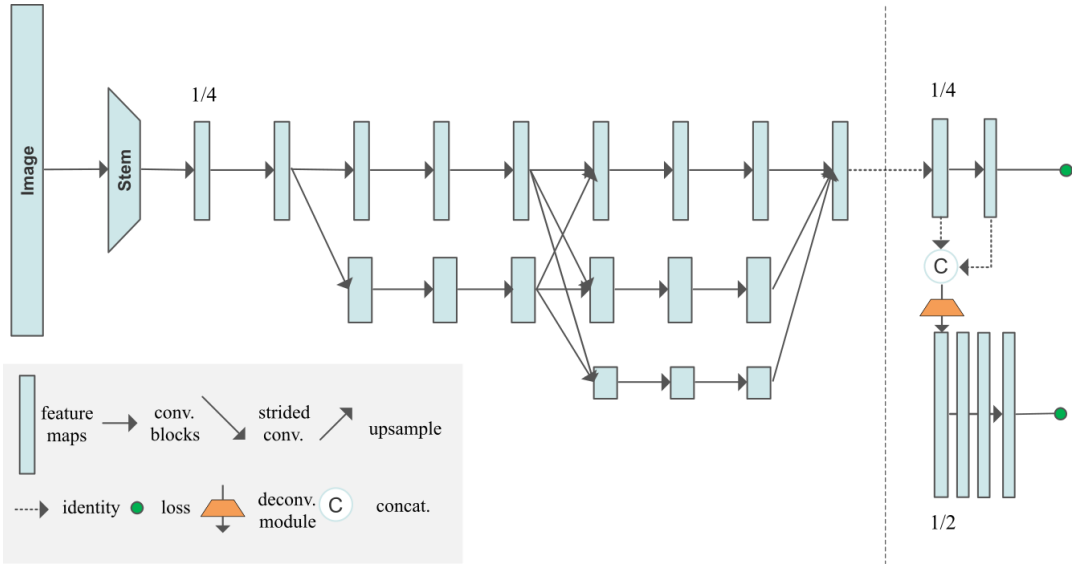

Fig. 3 An illustration of HigherHRNet [8]

3. SUMMARY

Current posture recognition algorithms are mainly used to process stationary objects, but in the field of human-computer interaction, there is a higher demand for posture recognition of moving human bodies. However, the moving human body has various postures and occlusion problems, which result in the identification of key points of the human body is not very accurate. Therefore, there is still a lot of room for development in the repair of limb occlusion.

Like a lot of visual tasks, human body posture estimating tasks so far seem to have fast into the performance bottleneck. But when combined with more practical application scenarios in the study, will find that there are still many questions are much harder to solve, this is the combination of production, is also the research community to the human body posture estimation field itself and the development opportunity.

In addition, the existing human posture recognition network still has the problem of large amount of computation and large number of parameters, and there is still a long way to go in reducing the number of parameters and the amount of computation.

REFERENCES

[1]. He K , Gkioxari G , P Doll ár, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017.

[2]. Ren S , He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[3]. Zhe C , Simon T , Wei S E , et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[4]. Wei S E , Ramakrishna V , Kanade T , et al. Convolutional Pose Machines[J]. IEEE, 2016.
[5]. Newell A , Huang Z , Deng J . Associative Embedding: End-to-End Learning for Joint Detection and Grouping[J]. 2016.
[6]. Zhou X , Wang D , P Krhenbühl. Objects as Points[J]. 2019.
[7]. Newell A , Yang K , Jia D . Stacked Hourglass Networks for Human Pose Estimation[C]// European Conference on Computer Vision. Springer International Publishing, 2016.
[8]. Cheng B , Xiao B , Wang J , et al. HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
[9]. Sun K , Xiao B , Liu D , et al. Deep High-Resolution Representation Learning for Human Pose Estimation[J]. arXiv e-prints, 2019.