ISSN 2616-5775 Vol. 6, Issue 4: 57-62, DOI: 10.25236/AJCIS.2023.060407

Research on Miners' Human Posture Detection Algorithm Based on MMPose

Tiantong Zhao

Department of Electrical and Electronic Engineering, School of Engineering, University of Manchester, Manchester, The UK
tiantongz@163.com

Abstract: *With the continuous improvement of intelligence in the coal industry, intelligent video analysis and judgment of dangerous behaviors that have occurred can no longer meet the needs of human posture detection in underground mines. In this paper, a human posture detection algorithm based on MMPose is proposed to determine the body posture of personnel under the mine at different times, in order to provide effective information for standardizing miner operations, disaster early warning, and accident rescue. The algorithm is based on MMPose technology, which improves and optimizes the existing shortcomings, facilitating the determination of the body posture of underground personnel at different times, and solving the low accuracy issues of occlusion image detection, multi-person image detection, and dark environment detection. The designed algorithm achieves labeling of 17 key points such as eyes, nose, ears, shoulders, elbows, hands, legs, knees, and feet, and can detect various postures such as standing upright, standing sideways, sitting, squatting, and walking. Experiments have shown that this algorithm has a high accuracy in human posture detection and prediction, and can better mark the key points of the human body for occluded images, multi-person images, and dark environments, meeting the real-time requirements of underground personnel posture detection.*

Keywords: *Human posture estimation; Machine vision; Key point detection*

1. Introduction

In recent years, with the continuous development of the coal industry, intelligent video analysis and judgment of dangerous behaviors that have occurred can no longer meet the needs of underground moving target monitoring. Therefore, the current posture detection of coal mine operators mainly provides effective information for disaster early warning and accident rescue to ensure timeliness. Research a human posture detection algorithm based on MMPose, which is convenient for judging the body posture of underground personnel at different times. Corresponding improvements have been made to address the problems existing in current human posture detection technologies, such as low accuracy for occlusion image detection, low accuracy for multi-person image detection, and low accuracy for dark environment detection.

In single person posture estimation [1-3], the problem of posture estimation is simplified by simply trying to estimate a single person's posture, and it is assumed that the person does not consider image content. Gkiox et al. used k-positions to jointly detect people and predict their posture and position. The final pose positioning is predicted by the weighted average of all active poselets. With the development of target detection and single person pose estimation, the two-step framework can further improve its performance.

There are two methods for estimating human posture, namely, machine learning based methods and inertial measurement based methods. Regarding machine learning based methods, L. Borkev [4] and others used state vector machines and tree based random forest algorithms to determine human posture; Zheng Lili et al. used RBF support vector machines to determine the posture of multiple human bodies; Huang Xinhan and others use the Fast R-CNN target detection algorithm and ZFNet [5] neural network to quickly identify targets and detect human posture, which can adapt to various environments and perform rapid detection; A real-time human posture estimation system based on convolutional neural networks was proposed by Xu Zhiqiang. This system uses convolutional posture machine algorithms to detect key points of the human body, and the 3D game engine Unity completes the construction of the virtual skeleton of the human body, achieving real-time human posture judgment. In terms of inertial measurement, Karagoz [6] and others have designed wearable devices using multi axis acceleration

ISSN 2616-5775 Vol. 6, Issue 4: 57-62, DOI: 10.25236/AJCIS.2023.060407

sensors, which can achieve estimation of human posture; Cao Yuzhen et al. proposed a human posture estimation algorithm based on MEMS accelerometers, mainly used to detect the posture of a human body falling.

Human posture estimation can be applied to the following scenarios [7-8]: 1) fall detection, fitness, dancing, ball games, martial arts sports guidance, and acupoint positioning; 2) Body language understanding: Airport and traffic police gesture translation, sign language translation; 3) Gait analysis, identification, abnormal movement identification; 4) Motion capture: 3D special effects movies; 5) Human-computer interaction: action control, gesture control; 6) VR, AR: Metacosmic digital human, Tiktok dance machine, 3D fitting, virtual anchor.

The above methods can achieve human posture estimation, but the effect is not ideal when applied in coal mines. The main problem is that the posture of underground personnel is relatively complex, and there are many problems such as occlusion, multiple people, and dark environment in the images collected by underground personnel, which can lead to low accuracy of the results. In response to existing problems, the MMPose algorithm is used to improve and optimize the shortcomings of existing algorithms, in order to better mark the key points of the human body for occluded images, multi-person images, and dark environments.

2. Human posture detection model

2.1. YOLO v3

Joseph Redmon et al. launched the YOLOv3 [9-10] version in 2018. The basic idea of YOLOv3 algorithm includes:

1) Label the area. The positive and negative samples are divided based on whether the candidate region is close to the real frame.

2) Using a convolutional neural network to extract image features and predict the category and location of candidate regions, a loss function can be established by comparing the network label value with the predicted value.

2.2. Darknet53

The backbone used for feature extraction in YOLOv3 is Darknet-53 [11], which draws on the network (Darknet-19) structure in YOLOv2. Unlike Darknet-19, Darknet-53 introduces a large number of residual structures, and uses Conv2D instead of Maxpooling2D with a step size of 2 and a convolution core size of 3×3. Through its classification performance on ImageNet, Darknet-53 has undergone the above modifications to greatly improve the network's operating speed while ensuring accuracy.

2.3. DeepPose

DeepPose [12] proposed at the CVPR conference in 2014 that it is the first time to use a deep convolutional neural network to achieve human posture estimation, and it is a milestone in the field of posture estimation. Through DeepPose, even without using complex representation knowledge modeling between joints or conducting interaction analysis between limbs, it is possible to use neural networks to learn human posture from a holistic perspective to achieve the most advanced effects.

2.4. OpenPose

OpenPose [13] is an open source library developed based on convolutional neural networks and supervised learning, which can achieve human posture estimation. It is a classic multi person attitude estimation method, which is a bottom-up algorithm.

2.5. MMPose

MMPose is a posture estimation and key point detection algorithm library of the OpenMMLab open source computer vision algorithm system, with built-in industry-leading 2D and 3D posture estimation algorithms for humans, faces, palms, and animals, which can detect the distribution of key points when the human body appears in any posture.

ISSN 2616-5775 Vol. 6, Issue 4: 57-62, DOI: 10.25236/AJCIS.2023.060407

3. Human posture detection algorithm based on MMPose

The open source algorithm library MMPose of the general visual framework OpenMMLab [14-15] for key point detection can perform 2D/3D key point detection on human bodies, faces, palms, and even animals, or call a camera for real-time Demo, call the pre training model in the MMPose library to predict images and videos, and train your own key point detection model on your own dataset. OpenMMLab is an internationally influential open source of artificial intelligence and computer vision led by China. The system covers many common tasks in computer vision, such as image classification and object detection. To sum up, MMPose has great advantages in human key point detection and has strong real-time performance. Therefore, it is possible to use human posture detection algorithms based on MMPose to conduct research on algorithms for human posture detection in underground mines.

3.1. Methods based on monocular images

Monocular images are easy to obtain and are not limited by the scene, but estimating 3D poses from 2D images is an ill-posed problem, where multiple different 3D poses can exist. Moreover, methods based on monocular images also face problems such as self-occlusion, object occlusion, and depth uncertainty. Due to the lack of 3D information, most current methods can only predict the root relative pose, which is a three-dimensional pose with the root joint as the coordinate origin.

Methods based on monocular images can be divided into direct prediction and 2D to 3D lifting depending on whether they rely on 2D HPE [16-17].

3.1.1. Direct prediction

Direct prediction does not rely on 2D HPE, but directly obtains 3D key point coordinates from image regression. The representative work C2F-Vol draws on the Hourglass network structure in 2D HPE and represents a 3D pose in the form of a 3D Heatmap. In order to reduce the huge storage consumption caused by three-dimensional data, a method of gradually improving the resolution in the depth dimension is adopted. Direct prediction can better utilize the information in the original image. The biggest drawback is that the mapping of 2D to 3D is a highly nonlinear problem, and the search range in 3D space is broader, making prediction difficult.

3.1.2. 2D-to-3D Lifting

Due to the high accuracy and generalization ability of 2D HPE, many methods choose 2D HPE [18] as an intermediate step to estimate 3D positions based on 2D positions (and original image features), and SimpleBaseline3D is one of the classic methods. This method takes 2D key point coordinates as input and directly maps 2D positions to 3D space through a fully connected layer connected by residuals. Although the model is very simple, the algorithm reached the SOTA level at that time, and experiments have proven that the error of most current 3D HPE algorithms mainly comes from the understanding of image information (2D HPE) rather than the 2D to 3D lifting process.

Since the above algorithm only uses 2D poses as input, it highly relies on the accuracy of 2D poses. If 2D HPE fails, it will seriously affect subsequent 2D to 3D lifting. To solve this problem, there are also some algorithms that learn 2D and 3D poses simultaneously, which can, on the one hand, introduce information from the original image for 2D to 3D lifting, and on the other hand, provide the possibility for mixed training of 2D/3D data sets, further improving the generalization ability of the algorithm.

The above is a single person pose estimation method. For 3D HPE in a multi person scene, similar to 2D situations, it can also be divided into top down and bottom up categories. The top down method requires first using a target detection algorithm to determine the bounding box of the human body, and then calculating the absolute coordinates of its root joint and the coordinates of other joints relative to the root joint for each bounding box of the human body. The bottom up method first predicts the position of all joints, and then connects the joints belonging to the same person based on the relative relationship between the joints to form a complete human body. The main advantage of the bottom up method is that its running time is basically not affected by the number of people to be detected, so the bottom up method has more advantages in crowded scenarios. The key point positioning of the bottom up method is completed based on the overall image, while the top down method is completed within the bounding box area. Therefore, the bottom-up method is more conducive to grasping global information, and can more accurately locate the absolute position of the human body in the camera coordinate system. The above features can highlight significant advantages in detecting human posture in underground mines.

ISSN 2616-5775 Vol. 6, Issue 4: 57-62, DOI: 10.25236/AJCIS.2023.060407

3.2. Methods based on multi eye images

Using multi eye images [19-20] for training can effectively solve the occlusion problem. In order to reconstruct 3D poses from multi eye images, the key lies in how to determine the positional relationship of the same point in the scene under different perspectives. Some methods introduce consistency constraints between multiple perspectives, such as simultaneously inputting images from two perspectives. For 2D pose inputs from one perspective, the 3D pose output from the other perspective is predicted based on the conversion relationship between the two perspectives. There are also some methods that use the triangulation method in stereo vision to aggregate 2D heat maps from all viewing angles to form a 3D volume, and then output the 3D heat map through a 3D convolutional network.

4. Experimental analysis

4.1. Experimental methods

This experiment mainly uses a method based on multi eye images to detect the human posture of underground personnel, which can effectively solve the occlusion problem and accurately mark key points in multi person images.

4.2. Experimental data set and key point selection

(1) Data set: The data set used in this experiment is independently constructed, mainly composed of multiple person images, but also contains some single person images, which are divided into two states: underground and underground. The dataset includes images of various postures of the human body, such as standing upright, standing sideways, sitting, squatting, walking, and other postures; Also selected are partially occluded images; In addition, the impact of dark environments on the algorithm is also considered to increase the accuracy of the algorithm and improve the generalization performance of the model.

(2) Key point selection: 17 key points were selected, including the left eye, right eye, nose, left ear, right ear, left shoulder, right shoulder, left elbow, right elbow, left hand, right hand, left leg, right leg, left knee, right knee, left foot, and right foot.

4.3. Experimental process

(1) Input an image of mine personnel, use VGG19 convolutional network to extract target features, output a set of feature maps, and then divide it into two branches Branch1&2, respectively using CNN network to extract partial confidence maps and partial correlation degrees;

(2) After obtaining the confidence and correlation degree, use the pair matching in graph theory to obtain partial association information and connect the human joint points. Due to the vectorial nature of the partial association, the pair matching in graph theory is relatively accurate, and finally, it is possible to generate an overall skeleton map of the human body;

(3) Further, PAF is used to obtain multi-person analysis results. The multi-person analysis results problem is essentially a graph problem. Therefore, information about the graph problem is obtained first, and then the Hungarian algorithm is used for calculation. The Hungarian algorithm is a combinatorial optimization algorithm that solves the task allocation problem in polynomial time. The goal of this algorithm is to find n 0 elements in different rows and columns in the transformation coefficient matrix to solve the optimal solution of the assignment problem.

4.4. Experimental results and analysis

To achieve a method based on human posture detection in coal mines, this experiment inputs multiple images related to mine personnel, with the input image shown in Figure 1 and the output image shown in Figure 2. As can be seen from Figure 1 and Figure 2, when inputting an image with multiple people in different poses, considering occlusion issues, multiple image issues, and dark environments, this algorithm can not only frame each person in the input image, but also detect and label the key points of each person. As shown in Figure 2, the positional relationship between the key points of the human body can be used to determine the miner's standing, walking, sitting, and other movement patterns. The different forms of multiple miners can be used to determine whether multiple miners are queuing up, or

ISSN 2616-5775 Vol. 6, Issue 4: 57-62, DOI: 10.25236/AJCIS.2023.060407

conducting activities such as survey or communication. The behavior of underground miners can be managed through analysis of forms and activities, and feature data can be provided for early warning of non-standard operation behaviors.

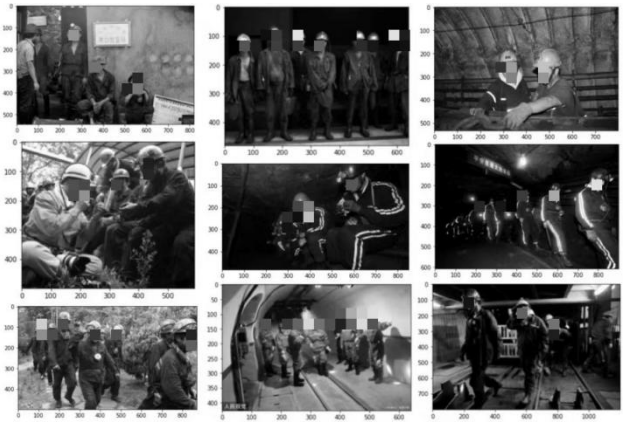

Figure 1: Input a picture

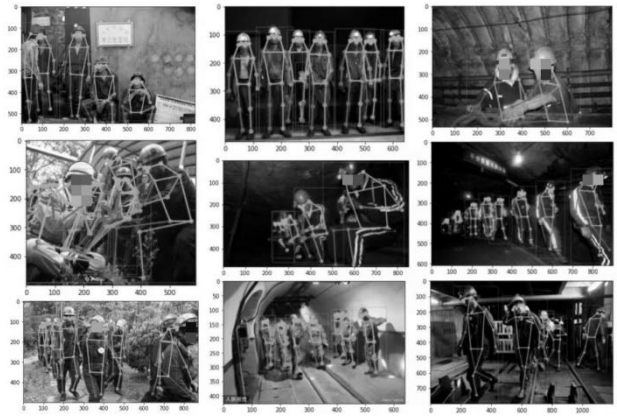

Figure 2: Input a picture

5. Summary

Research is conducted on better performance human posture detection algorithms. In the research, firstly, the current human posture detection algorithms are summarized, and the problem of applying existing algorithms to the underground environment of the mine is emphatically analyzed, which is reflected in the problem of low detection accuracy when there are problems such as occlusion, multiple people, and dark environment in the image. Based on previous experience and knowledge, a new pose detection algorithm based on MMPose technology is constructed, and methods based on monocular and binocular images are introduced. Finally, an experiment is conducted to verify the application effect of the algorithm. The final results show that the algorithm can accurately mark the key points of the human body when detecting the human posture of underground personnel, and can largely meet the basic needs of mine personnel posture detection. Therefore, it has high practicality and popularization.

References

[1] Liu Hao, Liu Haibin, Sun Yu, et al. Intelligent recognition system of unsafe behavior of underground coal miners [J]. Journal of China Coal Society, 2021, (S02): 1159-1169.
[2] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788. DOI: 10. 1109/ CVPR. 2016.91.
[3] Zheng L L, Huang X P, Liang R H. Human posture recognition method based on SVM[J]. Jounrnal of Zhejiang University of Technology, 2012, 40(6): 670-675, 691.
[4] He J Y. Teaching Research and Reform of Data Structure Course in Application-oriented Universities [J]. Computer Knowledge and Technology, 2021, 17(21): 108-110.

ISSN 2616-5775 Vol. 6, Issue 4: 57-62, DOI: 10.25236/AJCIS.2023.060407

[5] Ren S, 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.
[6] Kocabas, Muhammed and Karagoz, Salih and Akbas, et al. Multi Pose Net: Fast MultiPerson Pose Estimation using Pose Residual Network [C]. European Conference on Computer vision, 2018: 417-433.
[7] Chen C Q, Jiang L, Wang H. Gait prediction method of lower extremity exoskeleton based on SAE and LSTM neural network [J]. Computer Engineering and Applications, 2019, 55(12): 110-116.
[8] J. Redmon, A. Farhadi. YOLOv3: An incremental improvemen, 2018. Available from: https: // doi. org/ 10.48550/arXiv.1804.02767.
[9] Wei F F. Action Recognition Algorithm Based on Human Pose Estimation [D]. Graduate School of National University of Defense Technology, 2018.
[10] Cao Z, Hidalgo G, Simon T, Wei S E, Sheikh Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields[C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172-186. DOI: 10.1109/TPAMI.2019.2929257.
[11] Xu Z Q. Real-Time Human Poseure Estimation System Based On Deep Learning Study [D]. Yanshan University, 2019.
[12] Qian Z H, Gao L Q, Ye S. Method for detection of a student's pose in a multi-scene classroom based on meta-learning [J]. Jounrnal of Xidian University, 2021, 48(5): 58-67.
[13] Cao Y Z, Cai W C, Cheng Y. Body Posture Detection Technique Based on MEMS Acceleration Sensor [J]. Nanotechnology and Precision Engineering, 2010, 8(01): 37-41.
[14] Cai Z D, Ying N, Guo C S, et al. Research on multiperson pose estimation combined with YOLOv3 pruning model. Journal of Image and Graphics, 2021, 26(04): 0837-0846.
[15] Dixe Sandra et al. Optimized in-vehicle multi person human body pose detection [J]. Procedia Computer Science, 2022, 204: 479-487.
[16] Christoph Heindl et al. Large Area 3D Human Pose Detection Via Stereo Reconstruction in Panoramic Cameras, 2019. Available from: https://doi.org/10.48550/arXiv.1907.00534.
[17] Li X O, et al. Human Posture Detection Method Based on Wearable Devices. [J]. Journal of healthcare engineering, 2021, (4): 1-8.
[18] Lu Y K. Joint Human Parsing and Pose Estimation [D]. Dalian University of Technology, 2020.
[19] Sheng Y, Wang J Q. Research on Human Posture Recognition Based on Computer Vision [J]. Modern Information Technology, 2022, 6(16): 87-91+95.
[20] Bourdevl, Majis, Broxt, et al. Detecting people using mutually consistent poselet activations[C]. European Conference on Computer Vision, 2010: 168-181.