International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Human Action Recognition and Posture Prediction

M. Srividya[1], G. Swechha Reddy[2], S. Parameswari Devi[3], Sara Fatima[4]

[1]Assistant Professor, Matrusri Engineering College
[2,3,4]Information Technology Department, Matrusri Engineering College

Abstract:
Human action recognition and posture prediction aim to recognize and predict respectively the action and postures of persons in videos. They are both active research topics in computer vision community, which have attracted considerable attention from academia and industry. They are also the precondition for intelligent interaction and human-computer cooperation, and they help the machine perceive the external environment. In the past decade, tremendous progress has been made in the field, especially after the emergence of deep learning technologies. Hence, it is necessary to make a comprehensive review of recent developments. In this paper, firstly, we attempt to present the background, and then discuss research progresses.

Keywords: human action recognition; posture prediction; computer vision; human-computer cooperation; interactive cognition.

1. INTRODUCTION

The development of human society in recent years is known as the "AI Era", in which the development of intelligent technology needs self-learning and selfcognition abilities. The study of human action recognition and posture prediction enables machines to understand human behaviors and intentions and has been broadly applied in many fields. Research on human action has two basic topics: Human action recognition and posture prediction. Human action recognition involves detecting and classifying human actions from a time series (video frames, human skeleton sequences, etc.) that contains complete action execution, as shown. For example, the result of human body movement can be obtained by detecting the dynamic relationship between the static characteristics of the same frame and several adjacent frames. Human posture prediction automatically recognizes the current posture from temporally incomplete time series (video frames, human skeleton sequences, etc.). For example, self-driving vehicles can predict traffic police's actions, understand police's intentions, and make a judgment in advance.

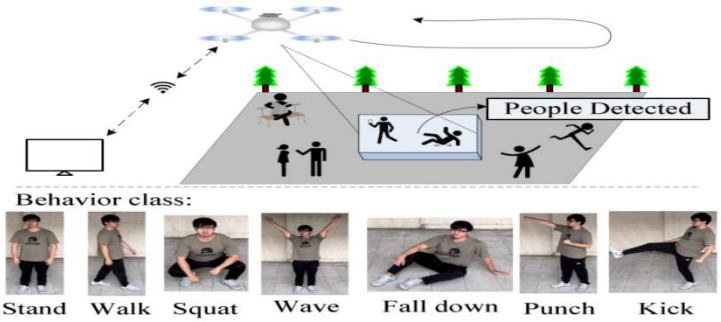


Fig.1: Example figure

The key difference between human action recognition and posture prediction is when making a judgment about an action[7]. Human action recognition is usually extrapolated from an entire video to an action tag. It is generally used in non-urgent scenarios, such as video surveillance and monitoring[8], and human action analysis[9–11]. Posture prediction is to infer the result before the action is completed, generally using to localize human body joint positions. For example, self-driving vehicles can predict pedestrian movements, conduct interactions between people and machines, understand people's intentions, and avoid dangerous accidents. It is typically used in application scenes with real-time requirements, such as human-vehicle interaction[12, 13], human parsing[14, 15], and human activity monitoring.

2. LITERATURE REVIEW

2.1 RMPE: Regional multi-person pose estimation:

Multi-person pose estimation in the wild is challenging. Although state-of-the-art human detectors have demonstrated good performance, small errors in localization and recognition are inevitable. These errors can cause failures for a single-person pose estimator (SPPE), especially for methods that solely depend on human detection results. In this paper, we propose a novel regional multi-person pose estimation (RMPE) framework to facilitate pose estimation in the presence of inaccurate human bounding boxes. Our framework consists of three components: Symmetric Spatial Transformer Network (SSTN), Parametric Pose Non-Maximum-Suppression (NMS), and Pose-Guided Proposals Generator (PGPG). Our method is able to handle inaccurate bounding boxes and redundant detections, allowing it to achieve a 17% increase in mAP over the state-of-the-art methods on the MPII (multi person) dataset.Our model and source codes are publicly available.

2.2 Extreme trust region policy optimization for active object recognition:

In this brief, we develop a deep reinforcement learning method to actively recognize objects by choosing a sequence of actions for an active camera that helps to discriminate between the objects. The method is realized using trust region policy optimization, in which the policy is realized by an extreme learning machine and, therefore, leads to efficient optimization algorithm. The experimental results on the publicly available data set show the advantages of the developed extreme trust region optimization method.

2.3 Survey of pedestrian action recognition techniques for autonomous driving:

The development of autonomous driving has brought with it requirements for intelligence, safety, and stability. One example of this is the need to construct effective forms of interactive cognition between pedestrians and vehicles in dynamic, complex, and uncertain environments. Pedestrian action detection is a form of interactive cognition that is fundamental to the success of autonomous driving technologies. Specifically, vehicles need to detect pedestrians, recognize their limb movements, and understand the meaning of their actions before making appropriate decisions in response.

3. METHODOLOGY

The key difference between human action recognition and posture prediction is when making a judgment about an action. Human action recognition is usually extrapolated from an entire video to an action tag. It is generally used in non-urgent scenarios, such as video surveillance and monitoring, and human action analysis. Posture prediction is to infer the result before the action is completed, generally using to localize human body joint positions. For example, self-driving vehicles can predict pedestrian movements, conduct interactions

between people and machines, understand people's intentions, and avoid dangerous accidents. It is typically used in application scenes with real-time requirements, such as human-vehicle interaction, human parsing, and human activity monitoring. As noted above, the problems of human action recognition and posture prediction are prevalent research topics. Nevertheless, there are still great challenges for researches.

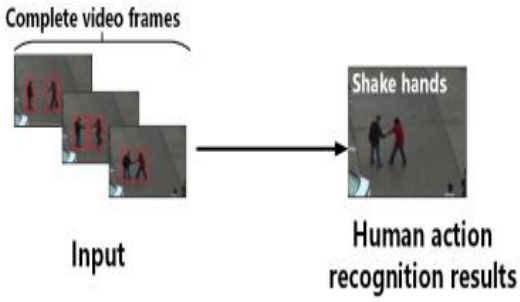
Disadvantages:

1. Large intra-class variation and inter-class similarity.
2. Complex scenarios lead to reduce accuracy
3. Long untrimmed sequences exist in many datasets.

Many relevant new ideas, frameworks, and approaches have been proposed in certain area. To better inspire future research and reveal the key trends of these fields, the study attempts to present the background, make a research overview and discuss progresses, datasets, various typical feature representation methods, and a variety of advanced human action recognition and posture prediction algorithms in recent years and other aspects. In addition, it is also pointed out that some future directions of human action recognition and posture prediction. The goal of this paper is to contribute to the field of computer vision, from theory, methodology, and system perspectives. It is believed that this survey can contribute to the field of computer vision, from theory, methodology, and system perspectives as well.

Advantages:

1. High accuracy.
2. There is no imbalance problem.



g

Fig.2: System architecture

4. MODULES:

In this project we have designed following modules

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Data augmentation: using this module to artificially increase the amount of data by generating new data points from existing data.
- Model generation: Building the model in colab - YOLOV5 - YOLOV6 - Mask-RCN. Algorithms accuracy calculated.
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

5. DESIGN SYSTEM

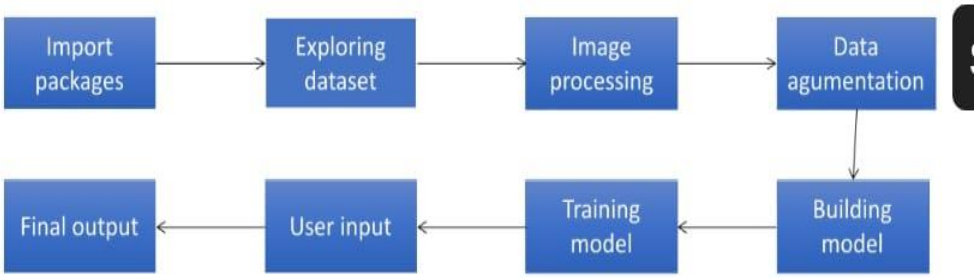


Fig.3: Design flow

The class diagram is used to refine the use case diagram and define a detailed design of the system. The class diagram classifies the actors defined in the use case diagram into a set of interrelated classes. The relationship or association between the classes can be either an "is-a" or "has-a" relationship. Each class in the class diagram may be capable of providing certain functionalities. These functionalities provided by the class are termed "methods" of the class. Apart from this, each class may have certain "attributes" that uniquely identify the class.

6. IMPLEMENTATION

YOLOV5- It is a novel convolutional neural network (CNN) that detects objects in real-time with great accuracy. This approach uses a single neural network to process the entire picture, then separates it into parts and predicts bounding boxes and probabilities for each component.

YOLOV6 – YOLOv6 is a single-stage object detection framework dedicated to industrial applications, with hardware-friendly efficient design and high performance. It outperforms YOLOv5 in detection accuracy and inference speed, making it the best OS version of YOLO architecture for production applications.

FasterRCNN – Faster R-CNN is a single-stage model that is trained end-to-end. It uses a novel region proposal network (RPN) for generating region proposals, which save time compared to traditional algorithms like Selective Search. It uses the ROI Pooling layer to extract a fixed-length feature vector from each region proposal.

7. RESULT

The result of the project ,is the action recognised by the model ,which has been trained .The action accuracy is also shown .The model is trained by inserting many frames of action by using deep learning algorithms.The model can learn several action at once . The result is shown below.

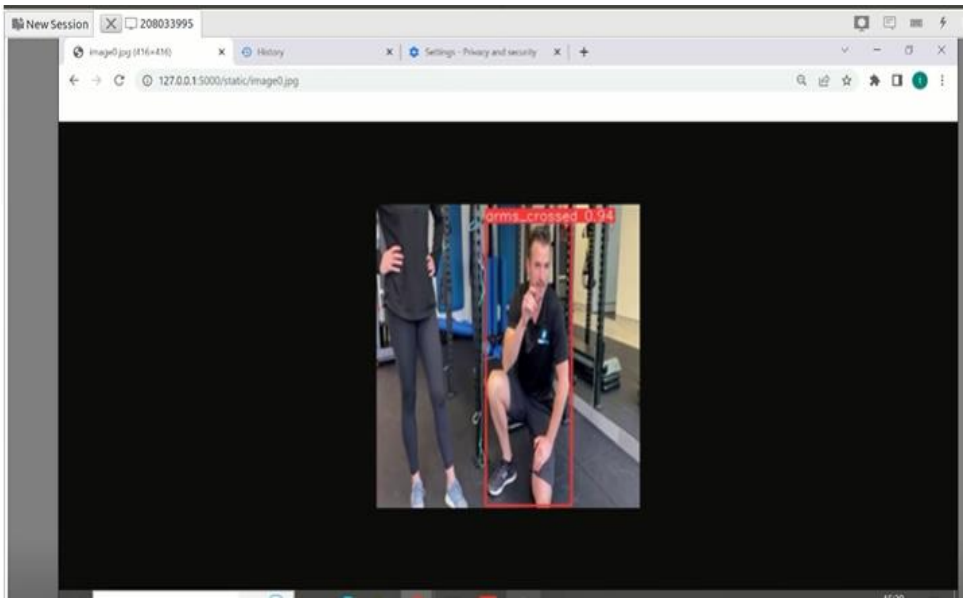


Fig.4: Prediction result

8. CONCLUSION

This literature review surveyed more than 200 papers related to human action recognition and posture prediction. Methods such as UDA, TPN, Action Genome, Sym-GNN[6] have been used in video understanding tasks, action analysis, and other relevant fields. Deep learning methods are improving, but differences still exist, such as two-stream adaptive graph convolutional network, Dynamic Directed Graph Convolutional Network, PoseC3D, and Channelwise Topology Refinement Graph Convolution Network, which outperformed state-of-the-art methods on the NTU RGB+D dataset.

9. REFERENCES

1. D. Y. Li, N. Ma, and Y. Gao, Future vehicles: Learnable wheeled robots, Sci. China Inf. Sci., vol. 63, no. 9, p. 193201, 2020.
2. H. S. Fang, S. Q. Xie, Y. W. Tai, and C. W. Lu, RMPE: Regional multi-person pose estimation, in Proc. 2017 IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 2353–2362.
3. H. P. Liu, Y. P. Wu, and F. C. Sun, Extreme trust region policy optimization for active object recognition, IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 6, pp. 2253– 2258, 2018.
4. L. Chen, N. Ma, P. Wang, J. H. Li, P. F. Wang, G. L. Pang, and X. J. Shi, Survey of pedestrian action recognition techniques for autonomous driving, Tsinghua Science and Technology, vol. 25, no. 4, pp. 458–470, 2020.
5. X. Y. Zhang, C. S. Li, H. C. Shi, X. B. Zhu, P. Li, and J. Dong, AdapNet: Adaptability decomposing encoderdecoder network for weakly supervised action recognition and localization, IEEE Trans. Neural Netw. Learn. Syst., doi: 10.1109/TNNLS.2019.2962815.