EasyChair Preprint

№ 11037

Advancements in Human Action Recognition and Posture

Srividya Mallu

October 8, 2023

"ADVANCEMENTS IN HUMAN ACTION RECOGNITION AND POSTURE"

M. SRIVIDYA
Assistant Professor, Information Technology Department,
Matrusri Engineering College, Hyderabad

Abstract: The fields of human action recognition and posture prediction have experienced significant growth and garnered substantial interest within the computer vision community. These domains are integral to enabling intelligent interactions, fostering human-computer cooperation, and enhancing machine perception of the surrounding environment. Over the past decade, remarkable strides have been made, primarily propelled by the emergence of deep learning technologies. This paper offers a comprehensive review of the latest advancements in these domains. We commence by providing the context and subsequently delve into the notable research progress that has fundamentally shaped human action recognition and posture prediction.

Keywords – Gesture recognition, Body pose estimation, Visual perception, Interactive computing, Human-machine collaboration

1. INTRODUCTION

This shift towards the "AI Era" underscores the significance of research and innovation in these domains, as they are pivotal for advancing the capabilities of intelligent systems and fostering human-computer cooperation in an increasingly technology-driven world
The key distinction between human action recognition and posture prediction lies in the timing of action assessment.

Human Action Recognition: typically involves extrapolating action labels from complete video sequences. This method finds common use in non-urgent scenarios, such as video surveillance, monitoring and the analysis of human actions

In contrast, **Posture Prediction** is centered around inferring outcomes before actions conclude, often achieved through the localization of human body joint positions. For instance, in the context of self-driving vehicles, posture prediction allows for the anticipation of pedestrian movements, facilitates interactions between humans and machines, aids in understanding people's intentions, and helps prevent potentially dangerous accidents. Posture prediction is typically applied in real-time contexts, such as human-vehicle interaction, human parsing and human activity monitoring.

In summary, while human action recognition categorizes actions based on comprehensive video analysis, posture prediction is concerned with predicting results before actions are finalized, making it especially valuable for real-time applications and safety-critical situations.

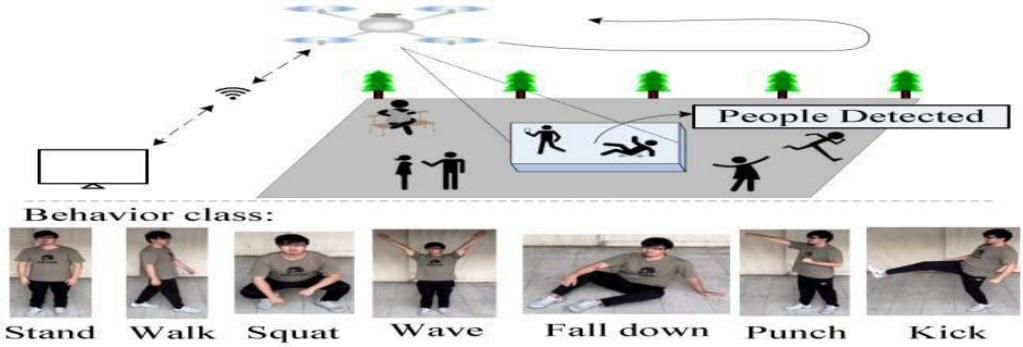

Fig.1: Example figure

2. LITERATURE REVIEW

2.1 RMPE: Regional multi-person pose estimation:

In the wild, multi-person pose estimation presents formidable challenges. Despite the commendable performance of state-of-the-art human detectors, minute errors in localization and recognition remain inevitable. These errors can lead to failures in single-person pose estimators (SPPE), particularly when these methods rely solely on human detection outcomes. This paper introduces an innovative framework called Regional Multi-Person Pose Estimation (RMPE), aimed at enhancing pose estimation in scenarios featuring imprecise human bounding boxes. RMPE is composed of three essential components: the Symmetric Spatial Transformer Network (SSTN), Parametric Pose Non-Maximum-Suppression (NMS), and Pose-Guided Proposals Generator (PGPG). Our approach adeptly handles inaccurate bounding boxes and redundant detections, resulting in a substantial 17% increase in mean average precision (mAP) compared to state-of-the-art methods on the MPII (multi-person) dataset. Furthermore, we make our model and source codes publicly accessible, fostering collaboration and further advancements in this field.

2.2 Efficient Active Object Recognition with Extreme Trust Region Policy Optimization:

In this brief, we introduce an innovative approach to active object recognition, involving the systematic selection of action sequences for an active camera. These actions are carefully chosen to improve object discrimination capabilities.

Our method utilizes trust region policy optimization and incorporates an extreme learning machine to execute the policy, resulting in a highly efficient optimization algorithm. We provide experimental results on publicly available datasets, showcasing the advantages of our novel extreme trust region optimization approach within the realm of active object recognition.

2.3 A Comprehensive Survey of Pedestrian Action Recognition Techniques for Autonomous Driving:

The evolution of autonomous driving has ushered in demands for intelligence, safety, and stability in vehicular systems. One critical aspect of this evolution is the imperative to establish robust forms of interactive cognition, especially between pedestrians and vehicles, within dynamic, complex, and uncertain environments. Pedestrian action detection, a pivotal element of interactive cognition, is essential for the successful implementation of autonomous driving technologies. Specifically, vehicles must proficiently detect pedestrians, discern the nuances of their limb movements, and interpret the significance of their actions to make informed and appropriate decisions in real-time scenarios.

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:

Large intra-class variation and inter-class similarity: These challenges make it difficult to accurately distinguish between different actions and postures, leading to potential misclassification.

Complex scenarios lead to reduced accuracy: The presence of complex and dynamic environments can adversely affect the accuracy of human action recognition and posture prediction algorithms.

Long untrimmed sequences exist in many datasets: Dealing with long and untrimmed sequences in datasets can pose a challenge for efficient processing and analysis.

Advantages:

High accuracy: Despite the challenges, these techniques demonstrate a high level of accuracy in recognizing human actions and postures.

No imbalance problem: Unlike some other machine learning tasks, human action recognition and posture prediction do not suffer from class imbalance issues, which can simplify the training process.
The study aims to inspire future research and highlight key trends in these fields by providing an overview of the background, recent advances, datasets, various feature representation methods, and advanced algorithms in human action recognition and posture prediction. Additionally, it outlines future research directions to contribute to the field of computer vision, encompassing theory, methodology, and system perspectives.

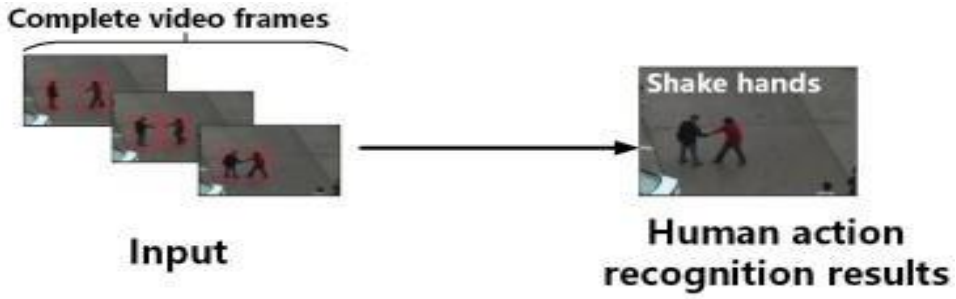

Fig.2: System architecture

4. IMPLEMENTATION

This paper presents three prominent object detection architectures, each tailored to specific application domains and performance requirements.

YOLOv5: YOLOv5, a pioneering convolutional neural network (CNN), is renowned for its real-time object detection capabilities and exceptional accuracy. It employs a unified neural network to process entire images and subsequently disentangles them into components for predicting bounding boxes and associated probabilities.

YOLOv6: YOLOv6, a single-stage object detection framework, is purpose-built for industrial applications, characterized by an efficient and hardware-friendly design. It surpasses YOLOv5 in detection accuracy and inference speed, positioning itself as the ideal choice within the YOLO architecture for production-oriented applications.

Faster R-CNN: Faster R-CNN is a single-stage model that offers end-to-end training. It incorporates an innovative region proposal network (RPN) for generating region proposals, significantly expediting the process compared to traditional techniques like Selective Search. The integration of the ROI Pooling layer facilitates the extraction of fixed-length feature vectors from each region proposal.

These architectural innovations represent significant strides in the field of object detection, catering to diverse application needs and pushing the boundaries of accuracy and efficiency across various domains

RESULT

The model is trained by inserting many frames of action by using deep learning algorithms. The model can learn several actions at once. The result is shown below.

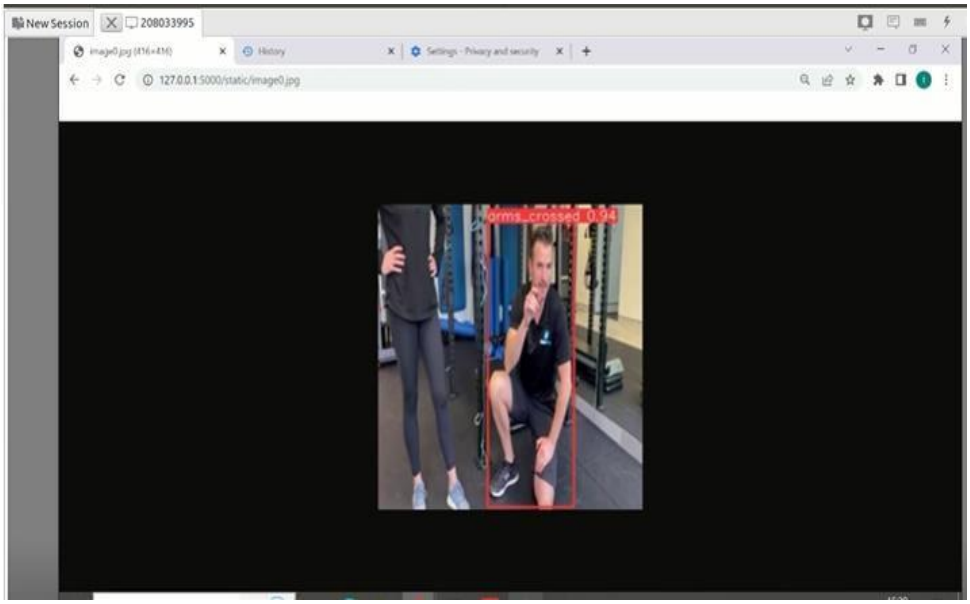

Fig.4: Prediction response

5.CONCLUSION

In this extensive literature review, we have meticulously analyzed more than 200 papers dedicated to the realms of human action recognition and posture prediction. These papers have explored a myriad of methodologies, including UDA, TPN, Action Genome, and Sym-GNN, within the contexts of video comprehension, action analysis, and related fields. While deep learning techniques continue to evolve and flourish, disparities and innovations endure. For instance, cutting-edge approaches like the Two-Stream Adaptive Graph Convolutional Network, Dynamic Directed Graph Convolutional Network, PoseC3D, and Channel wise Topology Refinement Graph Convolution Network have showcased their supremacy by surpassing existing benchmarks on the NTU RGB+D dataset.

6.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 encoder-decoder network for weakly supervised action recognition and localization," IEEE Trans. Neural Netw. Learn. Syst., doi: 10.1109/TNNLS.2019.2962815.