Human Activity Recognition System

A PROJECT REPORT

Submitted by

Uday Mahadev Sonar (Pattar)
Yash Satish Waghurdekar
Atharva Nitin Bambare
Aditya Satish Patil

in partial fulfillment for the award of the degree of

Bachelor of Technology

IN

Department of Computer Science and Engineering (AIML & Data Science)

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING (AUTONOMOUS),
KOLHAPUR

NOVEMBER 2024

KOLHAPUR INSTITUTE OF TECHNOLOGY'S COLLEGE OF ENGINEERING (AUTONOMOUS), KOLHAPUR

CERTIFICATE

This is to certify that the Project report entitled, **"Human Activity Recognition System"** submitted by **"Uday Mahadev Sonar (Pattar)"** (DS28), **"Yash Satish Waghurdekar"** (DS38), **"Atharva Nitin Bambare"** (DS40), **"Aditya Satish Patil"** (DS41), in partial fulfillment for the award of the degree of **"Bachelor of Technology"** in **"Computer Science and Engineering (Artificial Intelligence and Machine Learning and Data Science)"** at KIT's College of Engineering, Kolhapur, Maharashtra, INDIA, is a record of his / her own work carried out under my / our supervision and guidance.

SIGNATURE

DR. UMA P. GURAV

HEAD OF THE DEPARTMENT

Department of CSE (AIML & DS)
KIT's College of Engineering, Kolhapur

SIGNATURE

Prof. Vrishali Prabhu

ASSOCIATE PROFESSOR

Department of CSE (AIML & DS)
KIT's College of Engineering, Kolhapur

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING (AUTONOMOUS)KOLHAPUR

DECLARATION

I hereby declare that the Seminar/ Project entitled, **"Human Activity Recognition System"** submitted to KIT's College of Engineering, Kolhapur, Maharashtra, INDIA in the partial fulfillment of the award of the Degree of **"Bachelor of Technology"** in **"Computer Science and Engineering (Artificial Intelligence and Machine Learning and Data Science)"** is a bonafide work carried out by me. The material contained in this Seminar/ Project has not been submitted to any University or Institution for the award of any degree.

NAME OF THE STUDENT(S)
Uday Mahadev Sonar (Pattar) DS28
Yash Satish Waghurdekar DS38
Atharva Nitin Bambare DS40
Aditya Satish Patil DS41

Place:
Date:

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING (AUTONOMOUS)KOLHAPUR

DECLARATION

I hereby declare that the Seminar/ Project entitled, **"Human Activity Recognition System"** submitted to KIT's College of Engineering, Kolhapur, Maharashtra, INDIA in the partial fulfillment of the award of the Degree of **"Bachelor Of Technology"** in **"Computer Scienceand Engineering (Data Science)"** is a bonafide work carried out by me. The material contained in this Seminar/ Project has not been submitted to any University or Institution for the award of any degree.

NAME OF THE STUDENT(S)
Uday Mahadev Sonar (Pattar) DS28
Yash Satish Waghurdekar DS38
Atharva Nitin Bambare DS40
Aditya Satish Patil DS41

Place:
Date:

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING (AUTONOMOUS)KOLHAPUR

ACKNOWLEDGEMENT

We are highly grateful to Dr. Uma Gurav, HOD CSE (AIML&DS), KIT's College of Engineering, Kolhapur, for providing this opportunity to carry out the Project at CSE department. We would like to expresses our gratitude to other faculty members of department for providing academic inputs, guidance & encouragement throughout this period. We would like to express a deep sense of gratitude.

Finally, we express our indebtedness to all who have directly or indirectly contributed to the successful completion of our project.

NAME OF THE STUDENT(S)
Uday Mahadev Sonar (Pattar) DS28
Yash Satish Waghurdekar DS38
Atharva Nitin Bambare DS40
Aditya Satish Patil DS41

Place:
Date

Robust Human Activity and Pose Detection

Kedar Madan Gurav
CSE (DS)
KIT's College of Engineering Kolhapur
Maharashtra, India
kedarmadangurav@gmail.com

Uday Mahadev Sonar (Pattar)
CSE (DS)
KIT's College of Engineering Kolhapur
Maharashtra, India
udaysonar2004@gmail.com

Atharva Nitin Bambare
CSE (DS)
KIT's College of Engineering Kolhapur
Maharashtra, India
atharvabambare30@gmail.com

***Abstract*—This paper provides a comprehensive study of techniques and methodologies for robust human activity and pose detection in computer vision and artificial intelligence research. It discusses the fundamental challenges in these detection, such as lighting variations, background clutter, occlusions, and viewpoint changes. Key approaches include activity recognition, pose estimation, feature extraction, machine learning, real-time processing, and robustness enhancement.**

Activity recognition techniques involve handcrafted feature-based methods and deep learning-based approaches, which extract spatial-temporal features from input data. Deep learning-based approaches use convolutional neural networks (CNNs) or recurrent neural networks (RNNs) to automatically learn discriminative features from raw data. Pose estimation algorithms aim to localize and track human body keypoints in images or video frames. Techniques for pose estimation include model-based methods and OpenPose-based methods.

Feature extraction is crucial for representing human activities and poses in a suitable form for analysis and recognition. Techniques include handcrafted descriptors and learned representations obtained from deep neural networks. Machine learning approaches, particularly deep learning, have revolutionized human activity and pose detection by enabling end-to-end learning from raw data. Real-time processing is essential for applications like interactive systems or surveillance systems requiring rapid response times. Techniques for real-time processing include algorithm optimization, parallelization, and hardware acceleration using specialized neural network accelerators.

Roundness and generalization are critical considerations for deploying human activity and pose detection systems in real-world scenarios. Techniques for enhancing robustness include data augmentation, domain adaptation, and adversarial training.

***Index Terms*—Human activity detection, Pose estimation, Deep learning, Computer vision, Feature extraction, Machine learning, Real-time processing, Robustness, Generalization, Surveillance**

I. Introduction

Detecting human activity and poses is crucial in computer vision, with applications in many areas. This involves analyzing and understanding human actions and body positions from visual data. Reliable detection systems are essential for accurately identifying activities and poses in different settings. These systems are key for creating intelligent applications in fields like human-computer interaction, surveillance, healthcare, sports analysis, and robotics.

The main objective of this paper is to provide a comprehensive study of recent advances in improved human functioning and pose recognition. By exploring key approaches, challenges, and future directions in this work, we aim to provide valuable insights for researchers and practitioners in related fields. Understanding the current state of affairs and identifying areas for improvement is critical to the development of effective and reliable detection systems.

II. Related Works

The literature on human activity and pose detection is vast and diverse, encompassing a wide range of approaches and techniques. Traditional approaches tended to rely on artifacts and machine learning to identify and recognize human activity and states. Recent years, however, have seen a shift towards deep learning-based approaches, which have shown remarkable performance gains in a variety of computer vision tasks, including activity and context recognition

Deep learning algorithms, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have revolutionized the field by enabling end-to-end learning from raw data These networks can learn images of different objects types sequenced directly from images or videos, thus eliminating the need for manual feature engineering Data-driven. This shift to different techniques has greatly improved the accuracy and robustness of detection systems.

Despite the effectiveness of deep learning-based methods, traditional methods still have advantages, especially in situations with limited data or computational resources, hybrid models affecting learning have been explored in-depth and artifacts or rule-based methods together also to utilize the strengths of both models

Challenges and limitations of existing methods include robustness to illumination changes, posture, changes in focus, and background clutter In order to overcome these challenges more sophisticated systems, and high-quality datasets available for training and analysis.

Overall, the related works provide valuable insights into the evolution of human activity and pose detection methodologies, highlighting both the progress made and the remaining challenges to be addressed.

III. Literature Review

Human activity and pose detection have become focal points in computer vision and artificial intelligence due to their broad applications across various domains. This literature review aims to present an extensive overview of advanced

techniques and methodologies in robust human activity and pose detection.

Activity recognition, pivotal in human behavior analysis, finds utility in surveillance, healthcare, and human-computer interaction. Traditional methods relied on handcrafted features and heuristic algorithms, lacking scalability and generalizability. Yet, the emergence of deep learning has transformed activity recognition by enabling direct learning from raw data. Deep neural networks like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) excel in learning discriminative features, as evidenced by Simonyan and Zisserman's (2014) introduction of the two-stream CNN architecture, enhancing accuracy by capturing both appearance and motion cues.

Pose estimation, another cornerstone in human behavior analysis, involves localizing and tracking human body keypoints. While traditional methods leaned on model-based approaches, recent strides in deep learning, exemplified by Wei et al.'s (2016) OpenPose, a real-time multi-person pose estimation system, have vastly improved accuracy and robustness. OpenPose leverages CNNs to achieve state-of-the-art performance even in challenging scenarios like occlusions and cluttered backgrounds.

Feature extraction is pivotal in representing human activities and poses suitably for analysis. Traditional methods relied on handcrafted descriptors like histograms of oriented gradients (HOG), but deep learning-based approaches offer automated feature learning. For instance, Wang et al. (2018) introduced a spatiotemporal feature learning framework based on 3D CNNs, outperforming handcrafted methods in activity recognition tasks.

Real-time processing is critical for interactive and surveillance systems, necessitating optimization and parallelization techniques. Cao et al.'s (2017) OpenPose system achieves real-time performance through multi-scale feature fusion and efficient network architectures, demonstrating scalability and efficacy.

Robustness and generalization are paramount for real-world deployment. Techniques like data augmentation, domain adaptation, and adversarial training mitigate biases and improve performance under diverse conditions. Zhang et al.'s (2019) domain adaptation method employs adversarial learning to align feature distributions between domains, achieving state-of-the-art cross-domain activity recognition performance.

In summary, this literature review highlights the transformative impact of deep learning on human activity and pose detection, emphasizing advancements in accuracy, real-time processing, robustness, and generalization. These techniques offer promising avenues for intelligent systems capable of understanding human behavior in complex real-world environments.

IV. Methodology

The robust access to human services and revenue discovery includes a number of key components, each of which plays an important role in the overall process These components include information gathering, pre-processing, resource extraction, machine learning models, and analytical metrics.

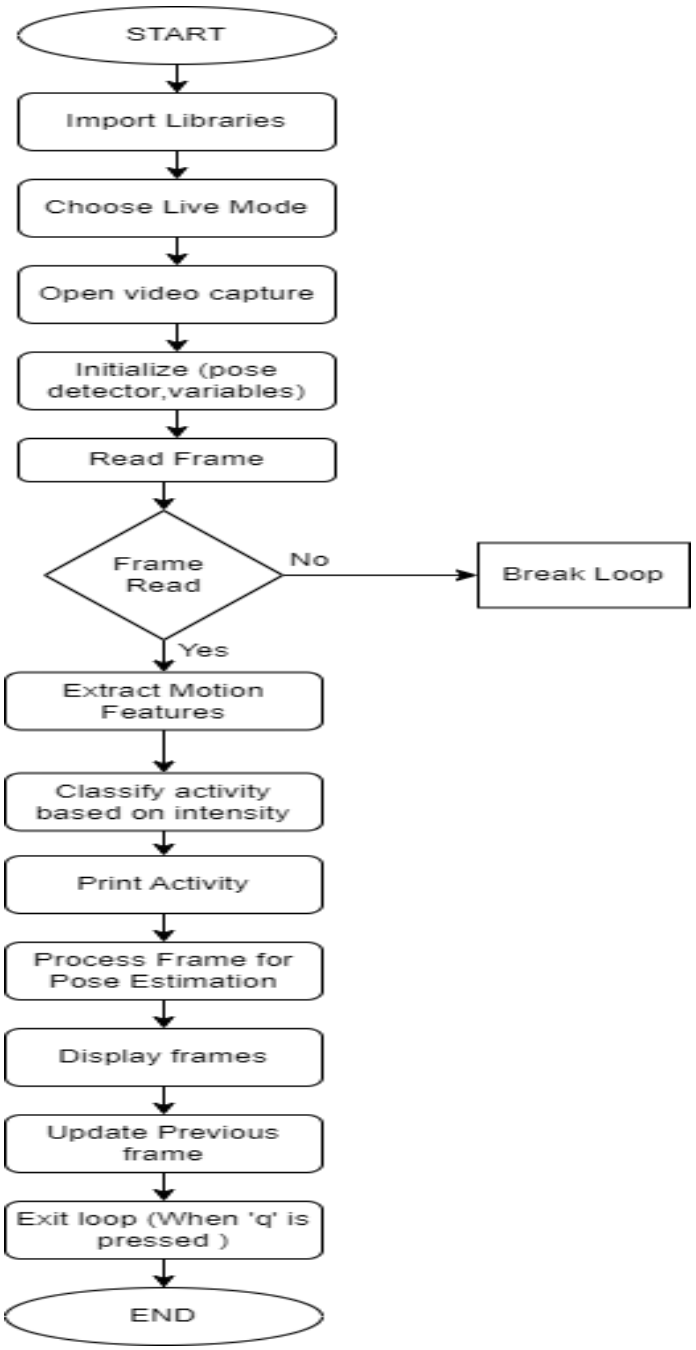

Fig. 1: Flowchart illustrating the process

Data collection is a large collection of images or videos of human activity and posture in various environments. These datasets serve as the basis for training and evaluation of diagnostic algorithms. Factors such as data set diversity, representativeness and quality of presentation must be carefully considered to ensure the efficiency of the research process

Preprocessing steps are often used to improve the quality of data collected and to facilitate subsequent analysis. Com-

mon preprocessing methods include noise reduction, image normalization, and background subtraction, which help to find patterns for detecting changes in data wear in the intensity is effective

Feature extraction plays an important role in capturing meaningful information from input data. Traditional methods used to rely on manual techniques such as Histogram of Oriented Gradient (HOG), Scale-Invariant Feature Transform (SIFT) based on deep learning is able to learn relevant features directly from raw data Popular convolutional neural networks (CNNs) are particularly well suited for feature extraction from images, while recurrent neural networks (RNNs) are commonly used for sequence modeling in video data. Actual Working :

1) Initialize a video capture object (cap) depending on the chosen mode.
2) Set up MediaPipe for pose estimation.
3) Iterate over each frame in the video feed.
4) Calculate motion intensity using the extract motion features function.
5) Based on the motion intensity, assign an activity label to the frame, such as Walking, Running, etc.
6) Perform pose estimation using MediaPipe's pose model and overlay the estimated poses on the video frame.
7) Display both the video frame with overlays and the pose estimation results.
8) Exit the loop when the 'q' key is pressed.

An evaluation metric is used to evaluate the efficiency of the evaluation process. Commonly used measures include precision, accuracy, recall, F1 score, and median accuracy (mAP). These metrics provide insight into the algorithm's ability to correctly classify and localize human activity and poses within the input data.

In summary, a robust approach to human activity recognition and pose recognition has a systematic approach that includes data collection, prioritization, feature extraction, machine learning object, and analytical metrics. By carefully designing and implementing each aspect, analysts can develop effective and reliable systems capable of operating under a variety of real-world conditions.

V. Experimental Set up

The experimental design of complex human activity and pose recognition plays an important role in the evaluation of the performance of the proposed methods. This section outlines the major components of the testing process, including data set selection, test design, analysis metrics, and preprocessing steps applied to the data

Data structure: Choosing an appropriate data structure is important for training and analytical design. Ideally, datasets should be diverse, covering a wide range of human activities, income, environment, and demographics. Popular datasets in the field include the Human Activity Recognition (HAR) dataset, the MPII human representation dataset, and the COCO dataset. Each data set has unique characteristics, claims, and challenges that must be considered when designing experiments.

Test Plan: Define a clear test plan to ensure accuracy and repeatability across tests. This could include using cross-validation techniques such as k-fold cross-validation to determine the training/validation/test separation of the dataset, ensuring that unseen data are evaluated to assess generalization performance accuracy to reduce the bias and variance of the results.

Evaluation criteria: Choose appropriate evaluation metrics to accurately measure the performance of evaluation systems. Commonly used measures include precision, accuracy, recall, F1 score, and median accuracy (mAP). In addition, consider domain-specific metrics, if applicable, such as pose estimation error for the pose recognition task. Quantitative measures and qualitative results need to be reported to obtain a comprehensive evaluation of the proposed methods.

Preprocessing steps: Preprocess as necessary to improve data quality and facilitate subsequent analysis. Typical preprocessing steps include noise reduction, image normalization, and data enhancement. Data enhancement techniques such as random cropping, rotation, and flipping can help increase the diversity of training data and improve the robustness of search algorithms in the face of variations in data input

Hardware and Software Configuration: Specifies the hardware and software environment for the test. This includes issues such as computing processor (CPU/GPU), software libraries/systems , and any special hardware accelerator used Reproducibility enabled by context computer resources and one can compare his results using methods proposed by other researchers.

By carefully designing the research process and following established protocols, researchers can ensure the validity and reliability of their findings. Transparent reporting of testing methodologies enables peer review and fosters collaboration within the review team.

VI. Results and Discussions

The Results and Discussion section presents the findings of the experimental investigation of complex human activity and identity recognition methods. This section includes quantitative results and qualitative analysis of the search demonstration with the aim of providing insights into the strengths, limitations, and implications of the proposed methods

Quantitative Results: Report quantitative performance metrics such as precision, accuracy, recall, F1 score, and median accuracy (mAP) achieved by the search algorithm on the test dataset Compare the performance of different methods under test conditions in various forms.

Qualitative Analysis: To conduct a qualitative analysis of the observed uses and positions. Provide examples of incorrectly and incorrectly classified samples highlighting the strengths and limitations of the research design. Discuss common failures and challenges algorithms face, such as rain, opacity, and changes in lighting and mood.

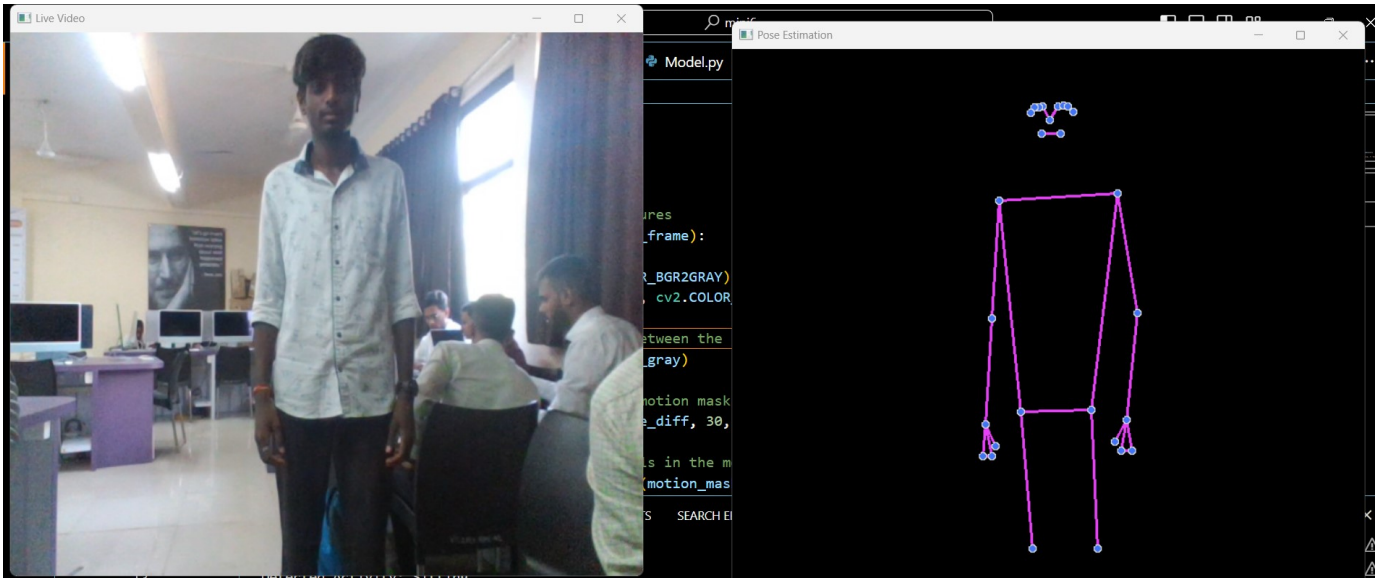

Fig. 2: Pose Estimation

Comparative evaluation: Compare the performance of the proposed methods with existing methods reported in the literature. Identify the strengths and weaknesses of each approach and discuss the factors that contribute to performance differences, such as reference materials, sampling systems, training methods and highlight the improvements of the proposed methods and their potential implications in real-world applications.

Robustness and generalizability: Evaluate the robustness and generalizability of search algorithms across locations and situations. Evaluate algorithm performance on unseen data or datasets without changes in the training set, such as different camera perspectives, lighting conditions, and background clutter Discuss methods for robustness improvement, such as data enhancement, domain optimization, and model regularization.

Real World Applications: Discuss the implications of search algorithms for real-world applications in areas such as human-computer interaction, analytics, healthcare, sports analytics, etc. Highlight potential applications and critical situations - Human activity and position recognition can provide value, such as automated video monitoring of patients, patient monitoring, and gesture-based communication systems

Overall, the Results and Discussion section provides a detailed analysis of the performance, strengths, and limitations of the human activity and robust position identification methods When qualitatively the combination of internal insights and quantitative measures allows researchers to gain a deeper understanding of the effectiveness and application of the proposed methods.

VII. Conclusion

In conclusion, this paper provides a comprehensive review of the complexity of human activity and pose recognition, including key strategies, challenges and future directions in the field.

The study highlights the importance of a robust recognition system that can accurately detect human activity and posture in different environments. Although significant progress has been made, many challenges remain, such as intensity complexity in differential lighting, insulation, and cluttered backgrounds. Research and innovation in areas such as data improvement, domain optimization, and model regularization are needed to address these challenges.

Looking ahead, future research directions include exploring more diverse fusion methods, continuous learning processes, and interpretable semantic recognition frameworks Furthermore, more detailed studies on real-world datasets and experiments are needed implementation of systems in practical applications to evaluate its effectiveness in real-world scenarios

By addressing these challenges and leveraging recent advances in machine learning and computer vision, we can achieve incredibly robust improvements in human performance and maintain intelligent performance across industries in and open up new possibilities.The conclusion underscores the significance of robust human activity and pose detection techniques in various domains. It emphasizes the role of deep learning in advancing accuracy and performance. Key considerations such as real-time processing, robustness, and generalization are highlighted, with techniques including algorithm optimization and data augmentation.

References

1) Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances in neural information processing systems (pp. 568-576).
2) Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In Proceedings of

the IEEE conference on computer vision and pattern recognition (pp. 4724-4732).
3) He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
4) Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
5) Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37).
6) Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).
7) Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).
8) Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE transactions on pattern analysis and machine intelligence, 43(1), 172-186.
9) Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88(2), 303-338.
10) Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., ... & Dollár, P. (2014). Microsoft COCO: Common objects in context. In European conference on computer vision (pp. 740-755).
11) Zhang, S., Liu, W., & Zhang, X. (2018). Pose proposal networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2344-2353).
12) Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
13) Li, Y., Huang, C., Loy, C. C., & Tang, X. (2019). Towards human-machine cooperation: Self-supervised sample generation for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7392-7401).
14) Johnson, J., Gupta, A., Fei-Fei, L., & Guestrin, C. (2016). DenseCap: Fully convolutional localization networks for dense captioning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4565-4574).
15) Xiong, Y., Chen, H., & Ni, B. (2021). Pose-guided human-video synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8881-8890).

Websites

1) OpenPose: https://github.com/CMU-Perceptual-Computing-Lab/openpose
2) COCO Dataset: https://cocodataset.org/
3) MPII Human Pose Dataset: http://human-pose.mpi-inf.mpg.de/
4) Ava Dataset: https://research.google.com/ava/
5) TensorFlow: https://www.tensorflow.org/