AI-ENABLED REAL-TIME PPE DETECTION USING DEEP LEARNING AND COMPUTER VISION

[1]Saritha Narayanan , [2]Akash Biju, [3]Yadukrishna KV, [4]Amaljith VR, [5]Fathimath Shameema TP

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student [1]Department of Computer Science and Engineering,
[1]St. Thomas College of Engineering and Technology, Kannur, Kerala, India.

Abstract : This system introduces an AI-enabled, sensor-free Personal Protective Equipment (PPE) monitoring framework aimed at enhancing real-time worker safety in industrial environments. By leveraging live video streams from existing surveillance and workplace cameras, the framework eliminates the need for wearable sensors while maintaining robust and non-intrusive monitoring. Advanced object detection models such as YOLOv8 and EfficientDet are employed to accurately detect PPE items including helmets, gloves, vests, and boots under diverse lighting and environmental conditions. To ensure proper PPE usage, human pose estimation techniques such as MediaPipe and OpenPose are integrated to analyze body posture and identify incorrect wearing or unsafe behaviors, including falls and collapse events. Furthermore, a 3D Convolutional Neural Network (3D-CNN) based activity recognition module classifies short video sequences to distinguish between wearing, carrying, or not using PPE. The fusion of object detection, pose analysis, and activity recognition enables reliable, scalable, and cost-effective real-time PPE compliance monitoring, making the system suitable for large-scale deployment in high-risk industrial environments.

IndexTerms **-AI, YOLOv8, Personal Protective Equipment (PPE), Computer Vision, EfficientDet, Workplace Safety.**

I. INTRODUCTION

Workplace safety is a critical concern in industrial and construction environments, where workers are frequently exposed to hazardous conditions. The use of Personal Protective Equipment (PPE), such as helmets, gloves, safety vests, boots, and goggles, plays a vital role in minimizing the risk of injuries and accidents. Regulatory bodies have established strict guidelines to ensure that workers wear appropriate safety gear at all times. However, enforcing these regulations remains a significant challenge, particularly in large-scale environments where continuous manual monitoring is impractical and prone to human error.

With the advancement of artificial intelligence and computer vision technologies, automated safety monitoring systems have gained considerable attention. Deep learning models, especially object detection algorithms, have shown promising results in identifying objects within images and video streams with high accuracy. Among these, YOLOv8 has emerged as an efficient and real-time object detection model capable of detecting multiple objects simultaneously with low latency. By leveraging such technologies, it is possible to develop intelligent systems that can automatically monitor PPE compliance without human intervention.

Despite these advancements, many existing systems face limitations such as reduced accuracy in complex environments, inability to verify proper PPE usage, and lack of real-time responsiveness. To address these challenges, this paper proposes an AI-based PPE detection system that utilizes deep learning techniques to monitor safety compliance in real time. The system is designed to detect multiple PPE items, analyze compliance, and generate alerts for violations, thereby providing an efficient, scalable, and reliable solution for enhancing workplace safety.

A. Problem Statement

Workplace safety remains a critical concern in industrial and construction environments, where workers are required to wear Personal Protective Equipment (PPE) such as helmets, gloves, safety vests, boots, and goggles. Despite strict safety regulations, ensuring proper PPE compliance is still a major challenge due to reliance on manual monitoring and supervision. Human-based inspection methods are often inefficient, time-consuming, and prone to errors, especially in large-scale or high-risk environments.

Existing automated systems for PPE detection are limited in terms of accuracy, scalability, and real-time performance. Many approaches focus only on detecting the presence of PPE items without verifying proper usage, while others fail to operate efficiently under varying lighting conditions, occlusions, and complex backgrounds. Additionally, most traditional safety systems lack the ability to provide instant alerts or continuous monitoring, reducing their effectiveness in preventing accidents.

Due to these limitations, there is a need for an intelligent, real-time, and automated system that can accurately detect PPE compliance and identify safety violations without human intervention. Such a system should be capable of processing live video streams, recognizing multiple PPE items simultaneously, and ensuring reliable performance in dynamic workplace environments.

II. LITERATURE SURVEY

A literature survey is a vital part of any research work as it provides insight into existing technologies, methodologies, and developments in the chosen domain. It helps identify strengths, limitations, and research gaps in prior studies, enabling researchers to build upon proven concepts while avoiding redundancy.

This survey provides a systematic review of several studies that have explored the use of computer vision and deep learning techniques for detecting safety equipment and ensuring compliance. These approaches primarily focus on object detection models, image processing techniques, and real-time monitoring systems to improve safety standards in industrial environments.

A. Vision-Based PPE Detection Using YOLO Architectures

Object detection models belonging to the YOLO (You Only Look Once) family have been extensively adopted for realtime PPE detection due to their computational efficiency and low inference latency. These models process input images in a single forward pass, enabling simultaneous localization and classification of PPE items such as helmets, reflective vests, gloves, and safety boots from CCTV or surveillance camera feeds. YOLO based methods are typically trained on labeled datasets containing PPE and nonPPE categories and perform well under controlled environmental conditions. However, their primary limitation lies in their focus on object presence alone. These approaches are generally unable to confirm whether detected PPE is being worn correctly. Moreover, detection accuracy may deteriorate in the presence of occlusions, poor illumination, motion blur, or complex industrial backgrounds, which are common in real-world settings.

B. Deep Learning–Driven Real-Time PPE Monitoring Frameworks

Beyond YOLO-based methods, several studies have explored deep learning frameworks built using platforms such as TensorFlow and PyTorch, incorporating detection architectures like YOLOv4, Single Shot MultiBox Detector (SSD), and Faster R-CNN. These systems enhance detection robustness through deeper feature extraction networks, multi-scale learning, and extensive data augmentation. Some frameworks also integrate object tracking mechanisms to maintain temporal consistency across frames. While such approaches demonstrate improved accuracy and scalability for real time PPE detection, they remain largely object centric. The absence of contextual understanding related to human posture and motion restricts their ability to identify unsafe behaviors such as incorrect PPE usage, falls, or abnormal body movements. Consequently, their applicability to comprehensive workplace safety assessment is limited.

C. PPE Compliance Verification Using Pose Estimation and Activity Recognition

More recent research has focused on combining object detection with human pose estimation and activity recognition to enhance PPE compliance assessment. Pose estimation frameworks such as MediaPipe and OpenPose extract skeletal keypoints to analyze body alignment, joint positions, and movement dynamics. When PPE detection results are spatially associated with these keypoints, systems can determine whether safety equipment is worn in the intended anatomical regions. Additionally, temporal modeling techniques, particularly 3D Convolutional Neural Networks (3D-CNNs), have been employed to capture motion patterns across video sequences. These models enable differentiation between actions such as wearing PPE correctly, carrying it without use, or completely neglecting it, and can also identify abnormal events including falls or collapse. Despite their effectiveness, the real-world deployment of such integrated systems remains constrained by high computational requirements and limited availability of annotated datasets for complex safety scenarios.

D. Identified Research Gaps and Limitations

An in-depth analysis of existing literature reveals several unresolved challenges. First, a clear detection-to-compliance gap exists, as most vision-based systems can identify PPE objects but fail to confirm correct usage. Second, a sensor–vision separation gap is observed, where solutions rely exclusively on either wearable sensors or visual data, rather than leveraging an integrated decisionmaking framework. Third, issues related to scalability and practical deployment persist, since sensorbased approaches increase cost and maintenance overhead, while vision-only systems often lack reliable real-time enforcement mechanisms. Additionally, many publicly available PPE datasets do not include explicit annotations for non-compliance scenarios such as missing helmets or vests, reducing their effectiveness for policy enforcement and safety analytics. Existing solutions are predominantly surveillance oriented and offer limited proactive intervention or immediate compliance feedback. To overcome these limitations, the proposed work introduces a unified, sensor-free PPE monitoring framework that combines deep learning based object detection, human pose estimation, and temporal activity recognition. This integrated approach transforms workplace safety monitoring from passive observation to intelligent compliance verification, enabling real-time decision-making while maintaining scalability, cost efficiency, and worker comfort.

III. METHODOLOGY

The proposed system adopts a real-time computer vision approach to monitor and ensure PPE compliance in workplace environments. The process begins with capturing live video input through a camera, such as a laptop webcam or surveillance system, which continuously provides frames for analysis. These frames are passed through a preprocessing stage where operations like resizing, normalization, and basic noise reduction are applied to enhance image quality and improve detection accuracy under different lighting and environmental conditions. Following preprocessing, a deep learning-based object detection model, YOLOv8, is employed to detect multiple PPE items including helmets, gloves, safety vests, boots, and goggles. The model processes each frame efficiently and generates bounding boxes along with class labels and confidence scores, enabling fast and accurate identification of safety equipment.

After detection, the system performs a compliance verification step to determine whether all necessary PPE items are properly worn by each individual. This ensures that the system not only detects objects but also evaluates safety adherence. If any required equipment is missing, the system flags it as a violation. The final output is displayed by overlaying bounding boxes and labels on the video frames for easy visualization. Additionally, the system can generate alerts or notifications when violations are detected, allowing for immediate corrective action. The entire pipeline is optimized for real-time performance, making it suitable for continuous monitoring in dynamic industrial environments while reducing the need for manual supervision.

IV. PROPOSED SYSTEM

A. Objectives

The primary objective of the proposed system is to develop an automated solution for real-time detection of Personal Protective Equipment (PPE) in workplace environments. The system aims to accurately identify multiple PPE items such as helmets, gloves, safety vests, boots, and goggles using deep learning techniques. Another key objective is to ensure PPE compliance by verifying whether workers are properly equipped and to generate alerts in case of violations. Additionally, the system focuses on achieving high accuracy, low latency, and scalability for practical deployment in industrial settings.

B. System Architecture

The proposed system follows a modular architecture consisting of three main layers: the Input Layer, Processing Layer, and Output Layer. The Input Layer captures live video streams through a camera and forwards the frames for analysis. The Processing Layer serves as the core component, where preprocessing and object detection using the YOLOv8 model are performed to identify PPE items. The Output Layer is responsible for displaying the detection results by overlaying bounding boxes and labels on the video frames and generating alerts for any safety violations. This layered architecture ensures efficient data flow and real-time performance.

C. Functional Modules

The system is divided into several functional modules to ensure clarity and scalability. The Video Capture Module is responsible for acquiring live video input from the camera. The Preprocessing Module enhances image quality through resizing and normalization techniques. The PPE Detection Module utilizes the YOLOv8 model to detect and classify PPE items in each frame. The Compliance Checking Module verifies whether all required safety equipment is present for each detected individual. Finally, the Alert and Visualization Module displays the results with bounding boxes and labels and generates alerts in case of missing PPE. These modules work together to provide an efficient and automated safety monitoring system.

D. Data Flow and Implementation

The data flow of the system begins with capturing live video frames from the camera, which are then passed to the preprocessing stage for enhancement. The processed frames are fed into the YOLOv8 model for PPE detection and classification. The detected outputs are analyzed by the compliance checking module to identify any safety violations. The final results are displayed in real time with visual annotations, and alerts are generated when necessary. The system is implemented using Python along with computer vision libraries such as OpenCV and deep learning frameworks supporting YOLOv8. The optimized implementation ensures realtime performance and efficient monitoring in dynamic environments.

V. PROPOSED SYSTEM DESIGN

The proposed AI-enabled PPE compliance monitoring system is designed using a modular and layered architecture to ensure real-time performance, scalability, and accurate safety monitoring in industrial environments. The system separates user interaction, data acquisition, processing, and alert generation into distinct layers, enabling efficient data flow and ease of maintenance. The overall architecture is designed to operate using live video streams and existing surveillance infrastructure.

1. Video Acquisition and Image Extraction

The system begins with the camera module, which captures real-time video streams from the workplace. These video streams are processed by the image extraction module, where relevant frames are selected for further analysis. This step ensures efficient handling of continuous video input while maintaining real-time performance.

2. Preprocessing Module

The extracted frames are passed to the preprocessing module, where image enhancement techniques such as resizing, normalization, and noise reduction are applied. This step ensures consistency in input data and improves the accuracy of the detection models under varying environmental conditions.

3. Dataset Integration Module

The system incorporates a dataset module that includes pre-collected images related to PPE usage and fall detection. These images are loaded and preprocessed similarly to real-time data. This module plays a crucial role in training and improving the robustness of the system by providing additional reference data.

4. Feature Extraction Module

In this stage, important visual features are extracted from both real-time and dataset images. These features include object shapes, textures, and motion patterns, which are essential for identifying PPE components and analyzing worker posture and activity.

5. Detection and Prediction Module

The extracted features are analyzed using deep learning models within the moderation and prediction modules. This combined stage identifies PPE items such as helmets, gloves, and vests, and determines whether workers are compliant or violating safety rules. It also detects unsafe conditions such as falls or improper equipment usage.

6. Alert Generation and Logging Module

If a violation is detected, the system triggers the alert generation module to notify supervisors in real time. At the same time, all detected events are stored in the database logging system for future analysis, reporting, and monitoring. This ensures both immediate response and long-term safety management.

Design Summary

The PPE compliance monitoring system adopts a modular and layered architecture to enable efficient real-time processing from video acquisition to alert generation. It integrates preprocessing, feature extraction, and deep learning-based detection to accurately identify PPE usage and unsafe activities such as falls. The system ensures high accuracy and low latency while maintaining scalability and ease of integration with existing surveillance infrastructure. Additionally, the inclusion of an alert mechanism and database logging supports immediate response and long-term safety analysis, making the system suitable for practical deployment in industrial environments.

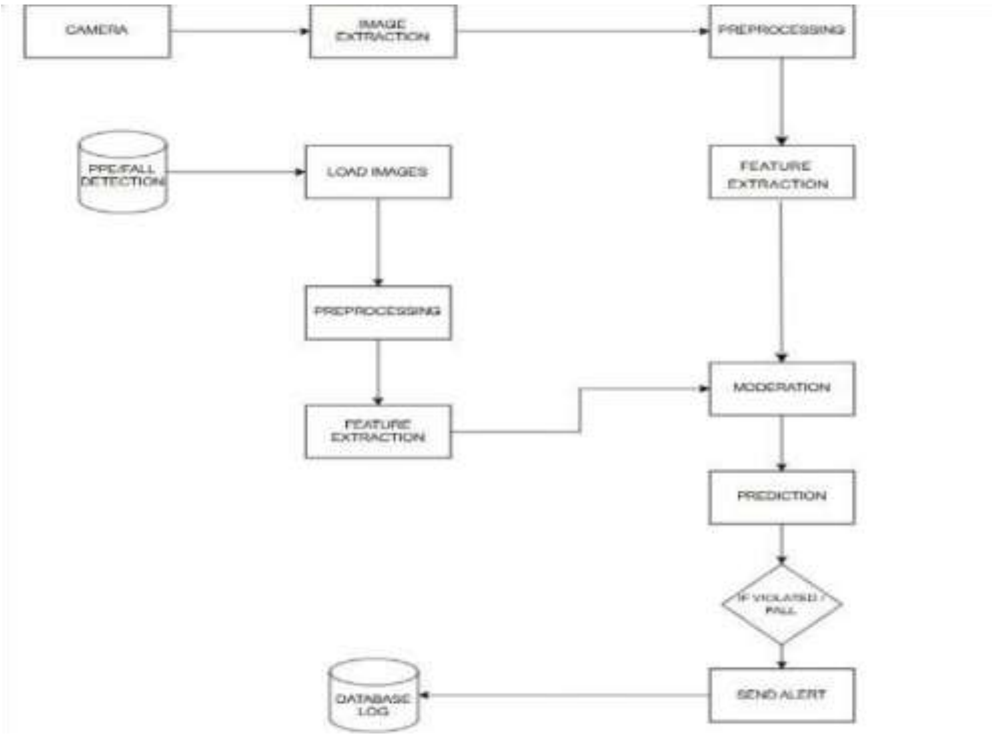

Figure 1. Architecture Diagram

The use case diagram illustrates the interaction between the worker, supervisor, and admin with the PPE compliance monitoring system. It shows how the system captures video, processes data, detects PPE usage, and identifies safety violations. The diagram also highlights alert generation and data storage functionalities for effective monitoring and management.

Actors

- **Worker**: The primary user who is continuously monitored by the system during work activities.
- **Supervisor:** Responsible for monitoring alerts and ensuring workplace safety compliance.
- **Admin**: Manages system configuration, datasets, and overall system operation.
- **System**: Represents the internal processing unit that performs detection, analysis, and alert generation.

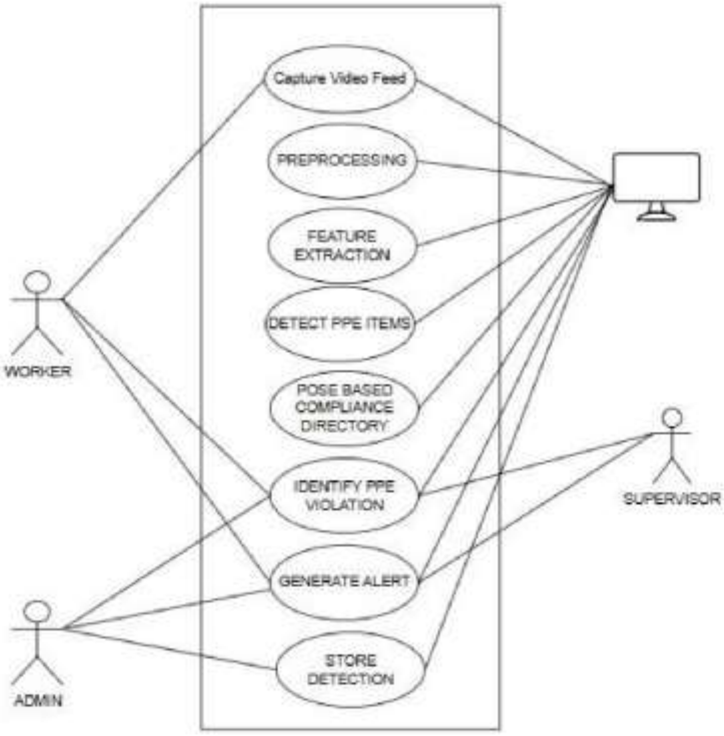

Figure 2. Use Case Diagram

1) **Capture Video Feed**
The system captures real-time video of workers through surveillance cameras for continuous monitoring.

2) **Preprocessing**
The captured video frames are processed to enhance image quality and prepare them for analysis.

3) **Feature Extraction**
Important visual features are extracted from the processed frames for further detection tasks.

4) **Detect PPE Items**
The system identifies PPE components such as helmets, gloves, and safety vests using deep learning models.

5) **Pose-Based Compliance Detection**
The system analyzes worker posture to verify whether PPE is worn correctly.

6) **Identify PPE Violation**
The system determines whether there is any safety violation, such as missing or improperly worn PPE.

7) **Generate Alert**
If a violation is detected, the system sends real-time alerts to the supervisor for immediate action.

8) **Store Detection**
All detection results and violations are stored in the database for monitoring and future analysis.

VI. RESULTS

The proposed PPE detection system was evaluated using a dataset containing images of workers with and without safety equipment to analyze its performance in real-time conditions. Key performance metrics such as detection accuracy, precision, and recall were observed during testing. The model demonstrated high accuracy in identifying PPE items including helmets, gloves, safety vests, boots, and goggles across different environmental conditions. The training and validation results indicate that the model effectively learns distinguishing features of each PPE category, achieving stable convergence with minimal overfitting.

In addition to quantitative evaluation, the system was tested on live video streams to assess real-time performance. The results show that the model is capable of detecting multiple PPE items simultaneously with low latency, making it suitable for practical deployment. The system successfully identifies missing safety equipment and generates appropriate alerts, ensuring continuous monitoring. Overall, the experimental results confirm that the proposed system is reliable, efficient, and capable of improving workplace safety through automated PPE compliance detection.

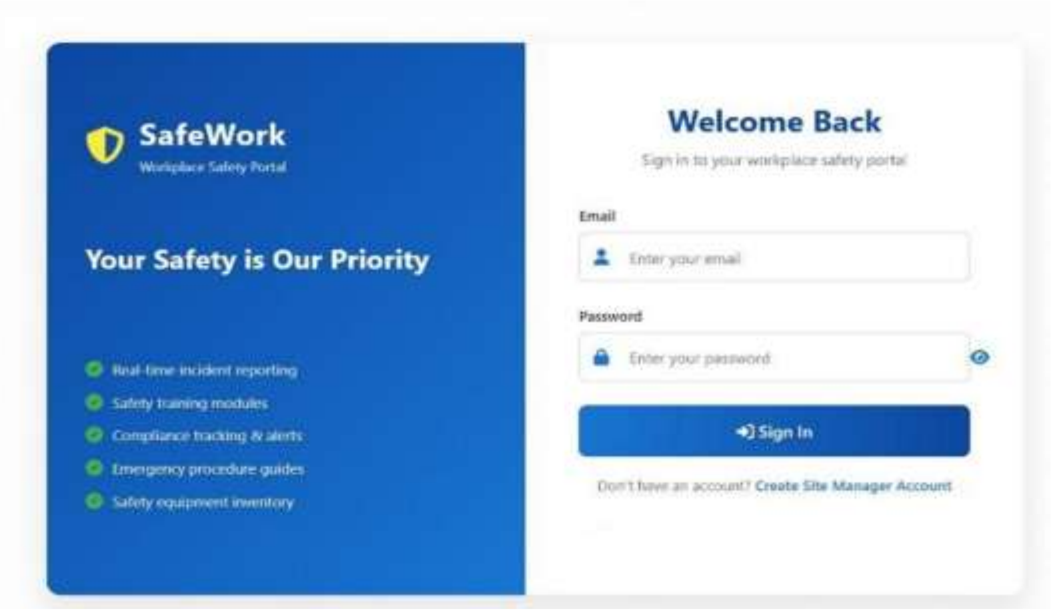

Fig 3. User Interface

Fig 4: PPE Detection

VII. CONCLUSION

This paper presented an AI-based real-time PPE detection system designed to improve safety compliance in industrial and construction environments. The proposed system utilizes advanced computer vision and deep learning techniques, specifically the YOLOv8 model, to accurately detect multiple PPE items such as helmets, gloves, safety vests, boots, and goggles from live video streams. By integrating real-time detection with compliance verification, the system ensures that workers adhere to safety regulations without the need for continuous human supervision.

The experimental results demonstrate that the system achieves high detection accuracy while maintaining low latency, making it suitable for real-world deployment. The modular and efficient design allows for scalability and easy integration with existing surveillance systems. Overall, the proposed solution provides a cost-effective, reliable, and automated approach to workplace safety monitoring, reducing risks and enhancing overall operational efficiency.

ACKNOWLEDGMENT

We would like to sincerely thank St. Thomas College of Engineering and Technology for their support. The resources provided through the Research and Development Lab were essential to the success of this project and gave us the environment we needed to turn this research into a working reality.

REFERENCES

[1] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” arXiv:1804.02767, 2018.

[2]] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” arXiv:2004.10934, 2020.

[3] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “YOLOv7: Trainable bagof-freebies sets new state-of-the-art for real-time object detectors,” arXiv:2207.02696, 2022.

[4] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv5,” GitHub repository, 2020–2024.

[5] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “YOLOX: Exceeding YOLO series in 2021,” arXiv:2107.08430, 2021.

[6] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in Proc. CVPR, 2020, pp. 10781–10790.

[7] K.He, X.Zhang, S.Ren, andJ.Sun, “Deepresidual learning for image recognition,” in Proc. CVPR, 2016, pp. 770–778.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll´ar, “Focal loss for dense object detection,” in Proc. ICCV, 2017, pp. 2980 – 2988.

[9] T.-Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common objects in context," in Proc. ECCV, 2014, pp. 740–755.

[10] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.

[11] H. Fan, B. Xiong, K. Mangalam, et al., "SlowFast networks for video recognition," in Proc. ICCV, 2019, pp. 6202–6211.

[12] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in Proc. NeurIPS, 2014, pp. 568–576.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. CVPR, 2005, pp. 886–893.

[15] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Proc. NeurIPS, 2017, pp. 5998–6008.

[16] R. Girshick, "Fast R-CNN," in Proc. ICCV, 2015, pp. 1440–1448.

[17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes (VOC) challenge," Int. J. Comput. Vis., vol. 88, pp. 303–338, 2010.

[18] K. He, G. Gkioxari, P. Doll´ar, and R. Girshick, "Mask R-CNN," in Proc. ICCV, 2017, pp. 2961–2969.