INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Recent Studies on Motion Detection for Education Environment using Computer Vision with Deep Architecture

Wan Muhammad Syafiq Wan Mithlan[1] & Hadhami Ab Ghani[2*]
Faculty of Data Science and Computing,
University of Malaysia, Kelantan (UMK), 16100, Pengkalan Chepa, Kelantan, Malaysia
Email: syafiq.m@umk.edu.my[1]; hadhrami.ag@umk.edu.my[2]

Submitted: 15/9/2025. Revised edition: 28/11/2025. Accepted: 5/1/2026. Published online: 28/7/2026
DOI: https://doi.org/10.11113/ijic.v16n1-2.685

Abstract—The integration of computer vision and deep learning within educational environments offers innovative methods for enhancing student engagement and focus assessment. Traditional classroom observation is often limited by time constraints and subjective biases, creating a need for objective and automated monitoring solutions. Recent advancements in AI-based motion detection, particularly using deep learning architectures such as CNNs, YOLO variants, R-FCN, and LSTMs, show promising results in tracking and analyzing student behaviors and engagement through movement and positioning. This review explores state-of-the-art models used in educational settings to classify and detect behaviors, emotional states, and engagement levels based on observable cues, including facial expressions, gestures, and posture. Furthermore, optimizing camera placement and using multi-angle configurations are identified as crucial factors for improving detection accuracy in complex classroom setups. Experimental findings on model performance, including precision, box loss, and classification accuracy, reveal these systems' efficacy and limitations, with some challenges persisting in precise localization. Despite these, the practical implications of real-time motion detection systems are significant, enabling educators to create responsive and interactive learning environments. This study underscores the potential of AI-enhanced monitoring in transforming classroom dynamics and promoting an engaging educational experience.

***Keywords*—Motion Detection, Deep Learning, Computer Vision, Student Environment, Classroom Monitoring**

I. INTRODUCTION

. All Online education has become a vital mode of learning for the younger generation in the post-pandemic era [1]. With the rise of digital tools and widespread internet access, online platforms now play a crucial role in delivering educational content and facilitating student-teacher interactions [2]. However, these platforms come with challenges, such as potential impacts on students' mental health, including stress from isolation and navigating online environments [3]. In contrast, traditional classrooms offer in-person interactions that support the development of social and professional skills like empathy, teamwork, and self-care [4]. Physical classrooms also promote vital communication skills, enabling students to engage more deeply in learning activities.

The integration of technology, especially artificial intelligence (AI), into education offers new ways to enhance engagement and focus [5]. One promising approach is using AI-based systems to monitor and analyze students' movements in educational group activities. Group activities are essential for cultivating teamwork, communication, and problem-solving abilities [6]. However, accurately assessing student participation and focus levels in group settings can be challenging. Manual observation methods are time-consuming and subject to observer bias, while recording lessons for review is often logistically complex [7]. Therefore, an AI-based motion detection system capable of identifying students' motions in real-time could provide educators with objective insights into participation levels and focus in group activities [8].

Deep learning architectures, particularly those leveraging motion detection systems, show promise in tracking and analyzing students' focus and engagement based on their movements and positions. By incorporating these technologies, educational settings could benefit from real-time data to tailor teaching strategies, making learning environments more dynamic and effective. Effective implementation of AI-based motion detection systems requires optimizing camera placement to maximize recognition accuracy. Factors such as camera angle, distance,

Advances in deep learning now allow for more accurate real-time detection and classification of human actions [11],

The integration of AI-based motion detection in classrooms holds significant potential for improving how student engagement is tracked, ultimately supporting more focused and interactive learning experiences.

This paper's major contributions are as follows:

- To study and identify students' motions in educational group activities using artificial intelligence-based approach.
- To study motion detection systems using deep architectures for detecting the students' focus in class based on the motions and positions of the student.

II. LITERATURE REVIEW

Nowadays, smart classrooms are very common among young learners. With the rise in college enrollment in recent times, there's a surge in the number of students attending classes, placing a growing burden on university educators to oversee classrooms and deliver lectures. As smart classrooms become more widespread, using deep learning to identify unusual student behaviors in class emerges as a vital step to enhance classroom productivity [12]. These modern learning spaces mark a big change in how knowledge is shared and learned. By using the latest developments in signal processing, internet tools, technology, and software, smart classrooms provide an exciting place for students to learn and interact. A smart classroom enables this knowledge distribution through technological advancements in signal processing, software web-based technologies, and hardware [13]. The implementation of smart classrooms as a pivotal national development initiative, the demand for monitoring and identifying student engagement in the classroom is on the rise. Extensive research within the education sector has focused on forecasting and assessing students' future academic performance and accomplishments based on their mobility and behavioral conduct. Historically, instructors have upheld order in the classroom and manually documented students' classroom conduct. This methodology is ineffective, fails to offer a comprehensive perspective of the classroom environment, and is susceptible to the subjective interpretation of the instructor [14].

2.2 Deep Learning for Motion

The last decade has witnessed a significant increase in interest in leveraging machine learning to create novel tools for educational assessment. Most of this research has centered on analyzing the movements and affective states of individual students [15]. Machine learning is a part of artificial intelligence that teaches machines to get better at

and coverage area are critical for accurate monitoring [10]. Research suggests that overhead cameras reduce occlusions and provide a broad field of view, while multiple cameras with overlapping fields can enhance tracking precision [9]. tasks by giving them data and algorithms. Artificial intelligence is a broad field where researchers strive to create smart systems that can make decisions independently. Machines achieve this by learning from data, recognizing patterns, and categorizing information to anticipate future outcomes and select the best solutions from many options available. Machine learning integrates various models and methods rooted in probability theory, statistics, data science, and advanced mathematics. These disciplines enable computers to make decisions. With sufficient training data, computers can discern patterns in numerical and visual data, along with the subsequent trends and behaviors they represent [16].

In recent years, there has been a growing fascination with comprehensive analysis of entire classrooms through video. In the realm of education, these approaches could provide an effective means of gauging student involvement, potentially reducing the reliance on manual assessment. Nonetheless, any machine learning technique designed to gauge underlying engagement levels during learning must rely on observable cues like head orientation, eye gaze, facial expressions, or body language [17].

III. REVIEW OF DEEP ARCHITECTURE FOR MOTION DETECTION IN EDUCATION

TABLE I. SUMMARY OF DEEP ARCHITECTURE APPROACH TO DETECT STUDENT MOTION

Deep Architecture	**Model**	**Summary**	**Reference**
General CNN algorithms	CNN-10	Study improves student behavior recognition with image skeleton data analysis.	(Zhou *et al.*, 2022)
	DCNN	Studies use deep neural networks for facial expression recognition with obscured faces.	(Hdioud B and Tirari M., 2023)
	RFCN	Optimized head detection enhances counting of small targets in classrooms.	(Liu *et al.*, 2020)
	Faster RCNN	Facial expression analysis improves classroom engagement detection using YOLO and Faster R-CNN, outperforming existing algorithms.	(Komagal & Yogameena, 2023)

Deep Architecture	Model	Summary	Reference
	3D CNN	C3D: Enhanced accuracy, robust performance, versatile, efficient spatiotemporal video analysis	(Xie *et al.*, 2021)
Specific CNN algorithms for deformed images	PIDM	PIDM enhances real-time emotion recognition on low-computing embedded devices for better classroom	(L. Li *et al.*, 2023)
	CFIN	CFIN uses context and attention to predict and reduce user dropout	(Feng *et al.*, 2019)
	ELDN	Ensemble deep learning network predicts MOOC dropout using integrated attention mechanisms.	(Kumar *et al.*, 2023)
	VBDTW	Voting-Based DTW enhances accuracy in non-invasive children's behavior detection.	(Wang *et al.*, 2023)
YOLO variants	Yolov4	Improved YOLO-v4 enhances classroom behavior recognition for teachers and students.	(Chen & Guan, 2022)
	Yolov5	YOLOv5 with SE attention improves multi-student learning behavior recognition.	(Trabelsi *et al.*, 2023), Wang *et al.*, 2022)
	Yolov8	Enhanced YOLOv8 improves classroom behavior detection accuracy and teaching efficiency.	(Chen *et al.*, 2023)
RNN	GRU	GRU selected for faster training and avoiding vanishing gradient issues.	(Botelho *et al.*, 2017)
ANN	ANN	ANNs optimize layers until output matches target; allows non-linear classification.	(Kumar, 2016)
LSTM	CDLSTM	CDLSTM model achieves 90% accuracy in detecting exam cheating	(Alsabhan, 2023)
	Online-LSTM, Offline-LSTM	Online-LSTM and Offline-LSTM improve action detection accuracy and efficiency.	(Carrara *et al.*, 2019)
	Eye movement analysis	Eye movement analysis algorithm classifies eye movement patterns efficiently.	(Xi, 2022)
	LSTM recurrent neural network	LSTM-based system achieves 98% accuracy in recognizing human motion data.	(Li, 2021)

Deep neural networks (DNNs) have sparked significant interest owing to their capacity for intricate data representation learning. This enthusiasm is especially evident in applications such as identifying motion among students in a classroom setting. A thorough review of prior research on the use of Deep Neural Network (DNN) architectures for detecting student motion has been conducted and is summarized in Table I. This table details the various deep architectures, models, and applications used in the studies, providing a concise summary of each, along with references to the original research. Additionally, the table includes information on the types of data employed, highlighting the scope and effectiveness of different methodologies across a range of applications.

In today's education, both traditional and digital platforms generate vast amounts of data. Analyzing this data helps educators enhance teaching and learning by identifying student behaviors and patterns. Educational data is a valuable resource that supports data-driven research and innovation. Deep learning, particularly Convolutional Neural Networks (CNNs), plays a key role in smart classrooms. CNNs can observe student behavior without disruption, automatically extracting image features for tasks like object detection. They reduce preprocessing by learning filters, improving image processing efficiency. Their strength lies in processing color images through 3D networks, making them efficient for image tasks [18], [39].

To detect the number of people in a classroom, models like R-FCN enhance Faster R-CNN by eliminating fully connected layers after RoIPooling and using position-sensitive feature maps to improve accuracy, especially in precise positioning [20]. A 3D CNN processes video data, capturing motion and temporal information, making it effective for behavior recognition and anomaly detection in videos [22].

The PIDM model efficiently analyzes classroom behavior using low-energy devices, providing instant feedback for teachers [23]. The CFIN model uses CNN-based context smoothing to predict dropout behavior, while the EDLN model combines Faster RCNN with time-series data to predict student status and identify at-risk students [24], [25]. The VB-DTW algorithm helps extract valid sensor data segments, and a 1D CNN algorithm can identify common classroom behaviors with 100% accuracy [26].

YOLO variants are widely used for real-time object detection in classrooms. YOLOv4, YOLOv5, and YOLOv8 detect student behaviors like hand-raising and head-turning. YOLOv5 integrates SE attention mechanisms for better multi-student behavior recognition [27], [28], and YOLOv8 enhances detection accuracy and efficiency for classroom activities [29], [30]. YOLO models process entire images at once, offering fast feedback and improving classroom management. They are robust in various environments, detecting even subtle student behavior changes.

RNNs and GRUs are used for motion detection and affective state recognition, as they capture temporal patterns in time-series data [31]. Artificial Neural Networks (ANNs) are efficient for recognition tasks, requiring less training and being suitable for complex problems, though challenges like overfitting remain [32]. LSTM networks handle sequential data, classify student behavior, and improve real-time predictions, making them ideal for detecting cheating and analyzing learning states [33], [34], [35]. The findings will also inform the improvement of AI-based applications in real-time monitoring systems, as illustrated in the overall framework in Fig. 1.

IV. THE PROPOSED METHOD FOR MOTION DETECTION

This chapter explains the research methods used. These methods help in solving the research problems mentioned. Chapter 3 discusses the Research Framework, Datasets and Summary.

The research framework consists of several stages, including a literature review, Phase 1, Phase 2, and the conclusion with future work. Each stage represents a crucial step in the problem-solving process, and each builds upon the previous one to ultimately achieve the research objectives. The literature review stage involves conducting an in-depth review and detailed analysis of existing studies related to motion detection, artificial intelligence, and swarm intelligence, which provides a strong theoretical foundation for the research. Phase 1 focuses on identifying the optimal camera placement, collecting motion data, and optimizing both parameters and hyperparameters to ensure that the data processing is both accurate and efficient. Phase 2 then transitions to the design and implementation of an updated motion detection algorithm, which is followed by the creation of a working prototype that can be tested and refined. Finally, the conclusion and future work section will review the key findings from the research and propose potential areas for further investigation, offering valuable insights that could guide future studies and technological advancements in the field. This structured approach ensures a comprehensive understanding of the problem while paving the way for innovative solutions. Additionally, the research will contribute significantly to the development of more efficient motion detection systems in educational settings. The findings will also inform the improvement of AI-based applications in real-time monitoring systems.

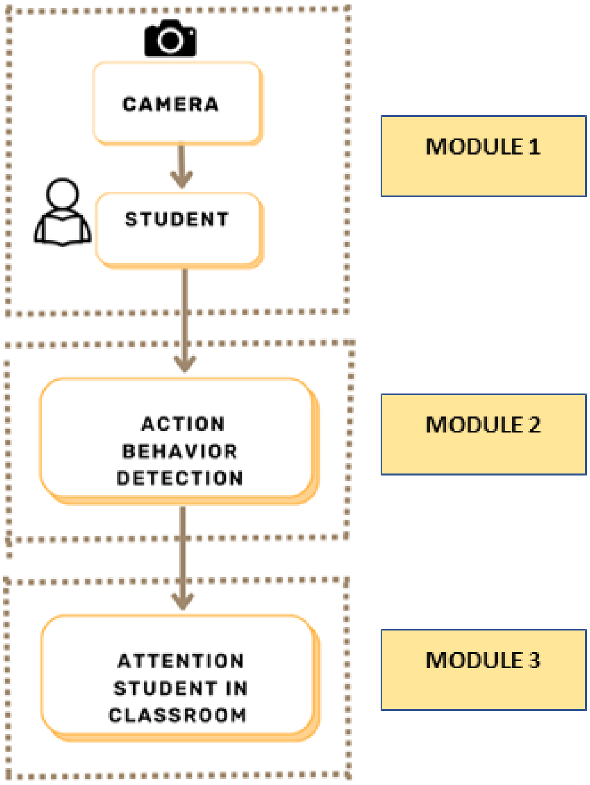

Fig. 1. Framework for Student Behavior Detection in the Classroom

A. The Research Framework

The research framework consists of a literature review, Phase 1, Phase 2, and conclusion with future work. Each step addresses a part of the problem-solving process. The literature review examines existing studies on motion detection, AI, and swarm intelligence. Phase 1 focuses on finding the optimal camera position, collecting motion data, and optimizing parameters and hyperparameters. Phase 2 involves designing and executing the updated motion detection algorithm with a prototype. The conclusion reviews the findings and suggests areas for future investigation.

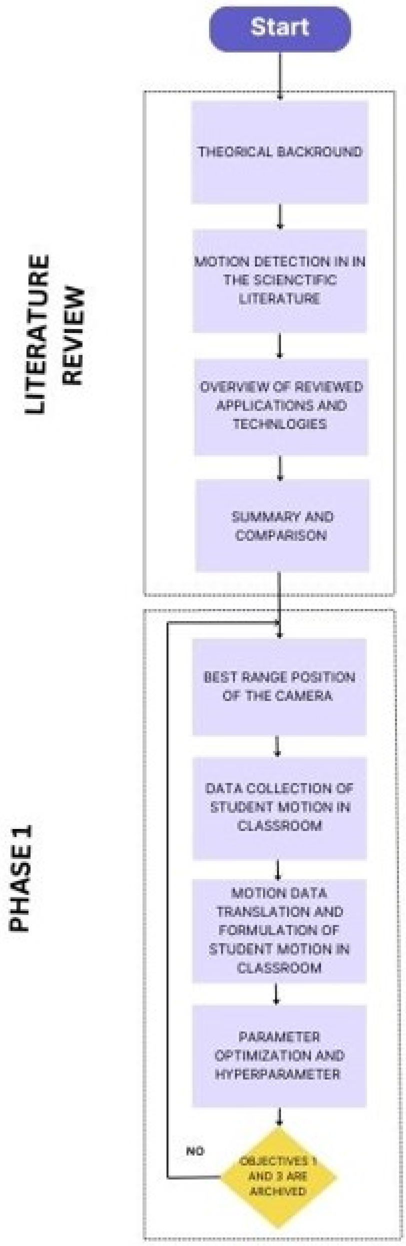

Fig. 2. Research Framework and Workflow for Motion Detection Flowchart

The research begins with a review of existing literature on motion detection in education, focusing on AI models. It evaluates the strengths and weaknesses of deep learning models like CNNs and RNNs. The study identifies relevant datasets and aims to develop a custom one for smart classrooms to address the lack of domain-specific data. Using advanced models like YOLO variants, this research aims to improve student behavior monitoring, contributing to better classroom management and learning outcomes. It also explores combining AI algorithms to improve motion detection systems for educational settings.

AI, particularly deep learning models like CNNs and RNNs, has transformed how student behaviors are monitored in classrooms. CNNs are effective at detecting static behaviors such as hand gestures and facial expressions, while RNNs track behavior changes over time. YOLO models, known for their real-time detection, are ideal for monitoring multiple behaviors simultaneously in crowded classrooms. YOLOv4, YOLOv5, and YOLOv8 are particularly fast and accurate, detecting subtle behaviors and providing immediate feedback to teachers. YOLO's ability to process large amounts of data efficiently makes it perfect for smart classrooms.

In Phase 1, the goal is to find the best camera placement and improve the motion data collection process to analyze student behavior accurately. The camera must cover the entire classroom to capture all student activities. This involves choosing the right camera angle, distance, and classroom size to ensure that all movements, both individual and group, are visible without disturbing the learning environment.

Motion data from video footage will be processed to identify key behaviors like hand-raising, head-turning, and body posture changes. This data will be converted into measurable features that reflect specific student behaviors, such as slouching or fidgeting, which may indicate disengagement. Deep learning models will categorize these behaviors into focused, distracted, or active participation.

The study will create a mathematical model of student behavior using motion data, quantifying actions like speed, frequency, and duration to predict behavior trends. For example, repetitive restless movements might suggest that a student is losing focus. This model provides real-time insights into student engagement.

The research will also optimize parameters and hyperparameters, such as the number of hidden layers, learning rate, batch size, and detection thresholds, to improve detection accuracy. This ensures precise behavior classification and reduces false positives, allowing for more reliable student engagement assessments. The conclusion reviews the findings and suggests areas for future investigation, which is summarized in the flowchart in Fig. 2.

B. Process Flow for YOLOv8-based Motion Detection in Classroom Videos

YOLO models, known for their real-time detection capabilities, are particularly effective in monitoring multiple behaviors simultaneously in crowded classrooms. Variants like YOLOv4, YOLOv5, and YOLOv8 offer high speed and accuracy, detecting subtle behaviors such as head-turning and standing, and enabling immediate feedback for teachers. YOLO's efficiency makes it the best choice for smart classrooms where timely insights are critical. YOLO stands out for its ability to process large amounts of data quickly and efficiently, making it the ideal model for real-time classroom monitoring. The complete workflow for this model development is outlined in Fig. 3.

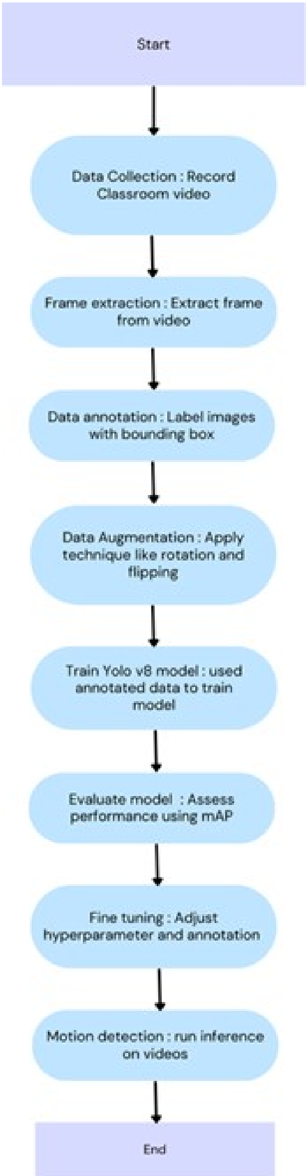

Fig. 3. Workflow for YOLOv8-based Motion Detection Model Development

In this phase, the flowchart outlines the key steps in implementing motion detection in a classroom using YOLOv8. It begins with data collection, where classroom videos are recorded, followed by frame extraction to convert video into individual images. Next, the images are manually annotated with bounding boxes around students or specific motions. After that, data augmentation techniques like rotation and flipping are applied to increase dataset diversity. The YOLOv8 model is then **trained** using the annotated data, and its performance is evaluated with metrics like mean Average Precision (mAP). Based on evaluation, the model undergoes fine-tuning by adjusting hyperparameters or refining annotations. Finally, the trained model is used for motion detection on classroom videos, completing the process

V. DATASET

In this paper, videos were sourced from YouTube, capturing various classroom scenarios. These videos were then labeled in Roboflow, where specific classes were created to categorize the behaviors. The classes include "Front-facing" for students looking directly at the camera, "Focus" for students who appear engaged in the lesson, "Talking" for students actively speaking, "Side-facing" for students looking away from the camera, and "Not Focused" for students showing signs of disengagement or distraction. These labeled categories were crucial for training AI models to detect and analyze student behaviors in the classroom. An example of this categorized dataset in the Roboflow platform is shown in Fig. 4.

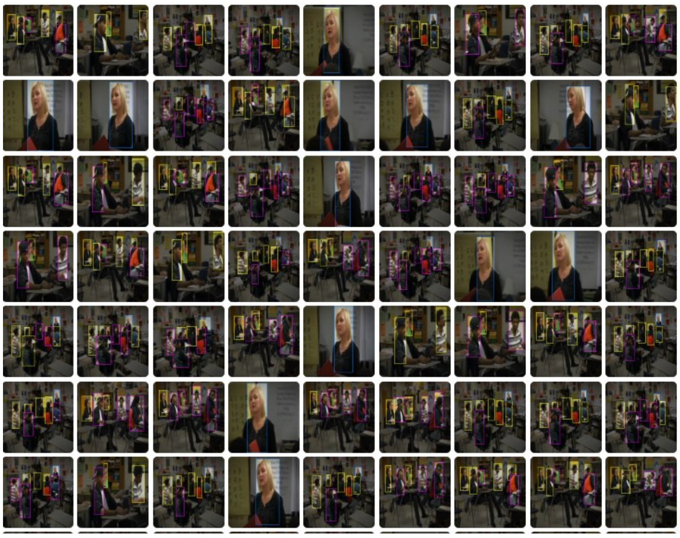

Fig. 4. Dataset on Roboflow

In recent years, computer vision technology has been increasingly used to automatically detect student behaviors in classrooms. However, the lack of publicly available datasets on student behavior has limited the application of video-based behavior detection in education. To address this, researchers have created several unpublished datasets, categorizing behaviors like listening, hand-raising, and writing. Some large-scale collections have also been developed from multiple schools, labeling behaviors such as standing and sleeping. Despite these efforts, many datasets cannot be shared due to privacy concerns related to real monitoring data (Yang Fan *et al*., 2024).

Effective human activity recognition (HAR) in educational settings requires specialized datasets, as highlighted by Sharma *et al*. (2022). Notable datasets include the Classroom Activity Dataset, which annotates behaviors like note-taking and hand-raising, and the Educational Interaction Dataset (EID), which captures student-teacher interactions. These resources are essential for developing AI-based motion detection models for educational applications, improving both academic and security outcomes.

However, domain-specific video datasets, especially those focused on classroom contexts, remain scarce. While image datasets like ImageNet are widely used, video datasets for real-time action recognition in classrooms, particularly for student-teacher interactions, are still limited. To address this, a new dataset was developed featuring classroom-specific actions compiled from YouTube and real classrooms, improving algorithm reliability in education research (Vijeta Sharma *et al*., 2021).

Creating high-quality datasets for detecting student movements is crucial for developing AI systems that enhance classroom management and student engagement. The process begins with selecting relevant YouTube videos showing various classroom scenarios, including disruptive behaviors. These videos capture a range of student actions, from active participation to boredom, representing different educational settings.

After selecting the videos, they are downloaded and converted into MP4 format. High-quality video ensures clear visibility of student and teacher movements for accurate AI model training. The videos are then broken down into frames using tools like OpenCV, which are categorized by specific behaviors, providing the foundation for training AI models.

The frames are uploaded to Roboflow, a platform for managing computer vision datasets. The images are categorized by their source video, and labeled with behaviors such as "Focusing" (active engagement), "Boring" (disengagement), or "Teaching" (teacher's role). These labels support AI model training to detect and analyze classroom behaviors.

To improve the dataset's diversity, data augmentation techniques like flipping, cropping, and rotating are applied, ensuring the model is trained on a variety of examples. The augmented dataset is then exported in formats compatible with AI frameworks like TensorFlow or PyTorch for training YOLO (You Only Look Once) variants, which excel at real-time object detection.

Refining the dataset further requires defining specific behaviors or motions to detect, such as sitting, standing, or hand-raising. Accurate annotations with keypoints or bounding boxes are essential for training the models. The dataset must also represent diverse classroom environments, capturing multiple angles to ensure broader applicability.s

Data annotation, whether manual or assisted by pre-trained models, is a critical step. Ethical considerations, including informed consent and anonymization, must be followed to comply with institutional review board (IRB) guidelines. Data augmentation helps reduce biases and class imbalance, improving the dataset's overall quality.

Once the dataset is prepared, it is split into training, validation, and test sets for optimizing AI model development. This structured process, from video sourcing to image augmentation, lays the foundation for AI systems capable of monitoring classroom dynamics in real-time. These systems provide valuable insights into student engagement, enabling teachers to make data-driven adjustments and foster more interactive and engaging learning environments.

VI. RESULTS

Roboflow provides comprehensive analytics to help users understand and optimize their computer vision projects. Their analytics likely cover multiple areas, including data preprocessing, model training, and deployment performance. During data preprocessing, Roboflow's augmentation tools enable users to enhance dataset diversity, and they track the frequency, type, and effectiveness of each augmentation to determine which variations improve model performance. For object detection and segmentation tasks, metrics such as Intersection over Union (IoU) and mean Average Precision (mAP) are likely used to assess model accuracy across various confidence thresholds.

During training, Roboflow tracks metrics like loss functions (e.g., cross-entropy or mean squared error for bounding boxes) and accuracy, providing insight into the model's learning progress over each epoch. For deployed models, they likely monitor inference performance metrics, including latency, GPU/CPU utilization, and memory usage, which are essential for real-time applications. They may also analyze user engagement within their app, using tools like Mixpanel or custom scripts, to improve user experience and optimize the interface based on user behavior. Additionally, Roboflow's analytics may include data pipeline performance metrics, covering aspects like data load times, dataset versioning, and integration capabilities, to streamline the dataset preparation and deployment process. This comprehensive approach to analytics helps users make data-driven decisions at every stage of model development and deployment.

eadings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

A. Dimension Insights

Fig. 5 provides an overview of the distribution of image sizes and aspect ratios within the dataset used for this study. The dataset contains a total of 300 images, all of which fall into the "large" size category, with dimensions that exceed the median values for both width and height.

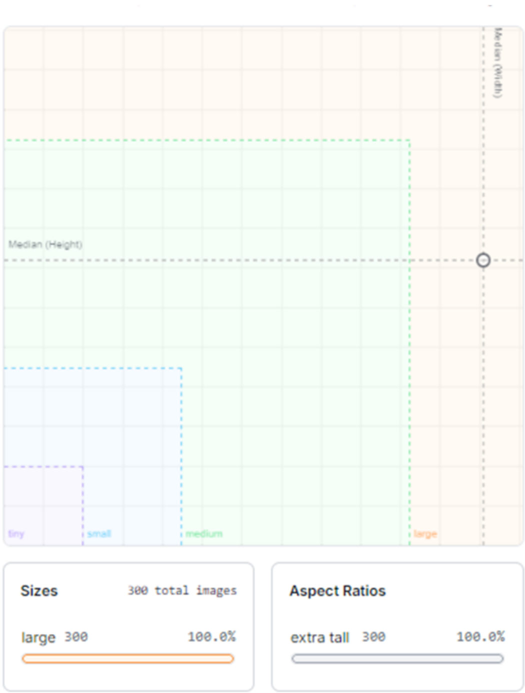

Fig. 5. Dimension Insights

The dashed lines on the graph represent the median dimensions for the dataset: 604 pixels in width and 360 pixels in height. All 300 images are classified as "extra tall" with a consistent aspect ratio that emphasizes vertical height, as evidenced by the histogram at the bottom right of the figure. This uniformity in both size and aspect ratio suggests that the dataset is well-suited for analyzing vertically-oriented objects or scenes, such as student postures in a classroom setting, where height-based behaviors (e.g., standing, sitting) are critical factors for analysis.

Such consistency in image size and aspect ratio ensures that the machine learning models, especially those based on object detection, will receive uniform input, thereby reducing the need for additional preprocessing steps like resizing or cropping, which can introduce noise or distortions. This is advantageous for the training process, as it allows the model to focus more on feature detection and behavior recognition, improving overall accuracy

B. Annotation Heat Map

The Annotation Heat Map (see Fig. 6) provides a visual representation of where annotations corresponding to the three classes focusing, boring, and teaching are located within the images in the dataset. The overlapping rectangles indicate the spatial distribution of the bounding boxes used to label these behaviors.

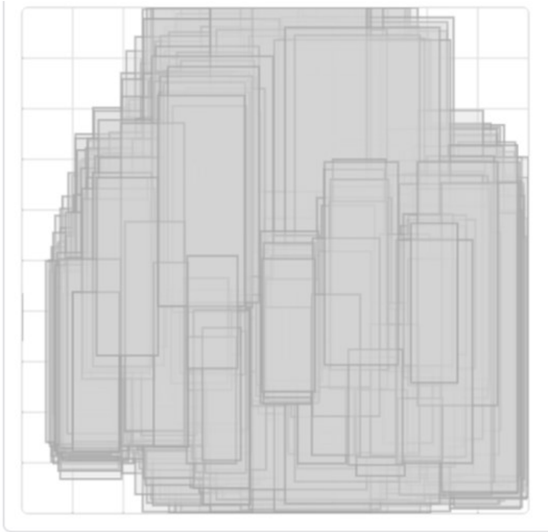

Fig. 6. Annotation Heat Map

From the heat map, it is evident that the annotations are concentrated towards the center of the images, suggesting that the majority of student behaviors being tracked, such as facial expressions or physical movements, occur in central regions of the classroom frames. This concentration may reflect the positioning of students in typical classroom layouts, where attention is often directed toward a central teaching area.

The variety in box sizes and overlapping regions indicates that the dataset captures a range of different behaviors and actions, spanning multiple scales and positions within the frame. This diversity ensures that the trained AI models will be exposed to a wide range of behavior types, improving the robustness of behavior detection across varying classroom conditions.

Furthermore, the heat map highlights the importance of carefully annotating key areas where student engagement (e.g., focusing or boredom) or teaching behaviors are prevalent. This balanced spread of annotations across different regions of the image contributes to comprehensive training of the AI models, allowing for more accurate behavior detection in real-world classroom scenarios

C. Histogram Object Count by Images

The dataset's annotation distribution is illustrated in a histogram (Fig. 7), showing the number of object classes annotated per image. The majority of images (155) are annotated with six to seven object classes, reflecting a high level of complexity and diversity in these images. A smaller portion of the dataset, 59 images, contains two to three class annotations, while 49 images are labelled with just one class. There is a noticeable drop in frequency for images annotated with four to five classes, with only 38 images falling into this category. The dataset also includes very few images (approximately one or two) with eight to nine annotated classes.The higher frequency of images with several object annotations can potentially improve model robustness in detecting and classifying diverse objects.

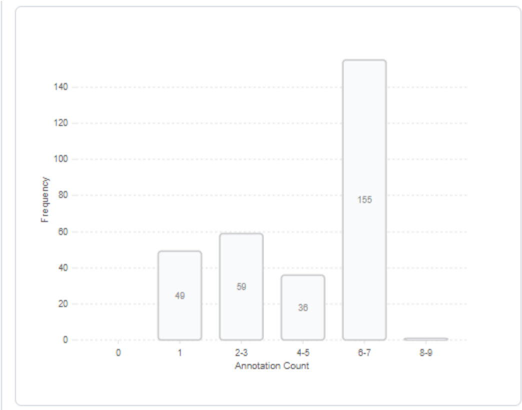

Fig.7. Histogram Object Count by Images

D. Mean Average Precision

The training graph in Fig. 8 illustrates the Mean Average Precision (mAP) performance during the development of the AI model for detecting student behaviors.

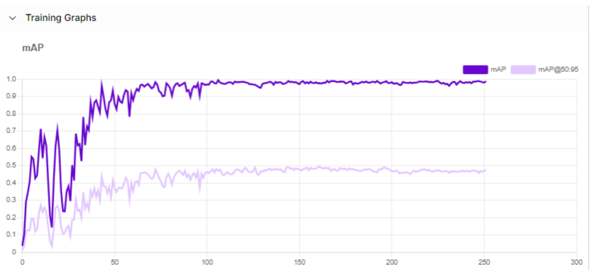

Fig. 8. Mean Average Precision

The graph presents two key lines which are one representing the mAP at an Intersection over Union (IoU) threshold of 0.5 (dark purple) and the other indicating the mAP over a range of IoU thresholds from 0.5 to 0.95 (light purple). The dark purple line demonstrates the model's rapid improvement in early epochs, with a sharp rise in mAP values, indicating that the model quickly learns to identify and classify behaviors. Around the 50th to 100th epoch, the mAP stabilizes and approaches a value near 1.0, suggesting the model has attained a high level of accuracy in object detection at this standard IoU threshold. Conversely, the light purple line, which evaluates performance over a stricter range of IoU thresholds, shows a more modest mAP, stabilizing around 0.6 to 0.7. This gap reflects the challenge of achieving precise localization in object detection, where higher IoU thresholds demand more exact overlap between predicted and ground truth bounding boxes. Despite this, the model performs consistently well, though there is room for improvement in precise localization. Overall, the graph confirms that the model has converged effectively, showing robust detection capabilities, with slight potential for refinement at higher IoU thresholds to further enhance performance in more demanding scenarios.

E. Box Loss

The graph of box loss over 300 training epochs as shown in Fig. 9 provides valuable insight into the performance of the object detection model. Initially, the box loss is relatively high, approximately 2.6, which reflects the inaccuracy in the model's early bounding box predictions. However, as the training progresses, there is a rapid reduction in loss during the first 50 epochs, indicating that the model is learning and refining its ability to predict bounding boxes more accurately. After 100 epochs, the rate of improvement slows, and the loss begins to stabilize, plateauing between 1.75 and 1.85. This plateau suggests that the model has reached convergence, with further training resulting in only minor improvements. Small fluctuations observed in the latter stages of training are typical and indicate a stable trend, suggesting the model is performing consistently. Overall, the decreasing box loss demonstrates the model's effective learning process, while the eventual stabilization reflects the attainment of optimal performance in predicting bounding boxes.

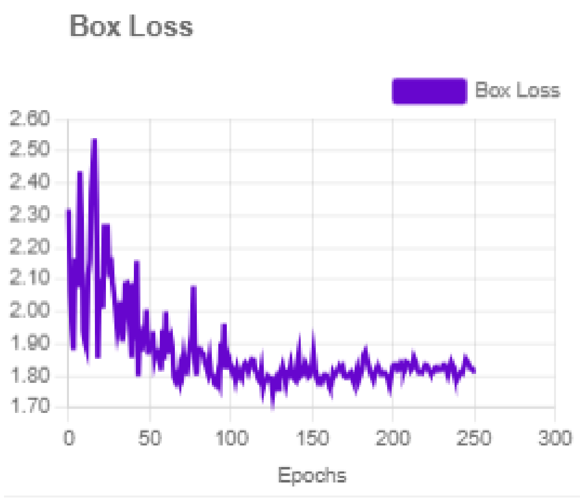

Fig. 9. Box Loss

F. Class Loss

The graph in Fig. 10 shows the trend of 'Class Loss' across 300 training epochs, which measures how well the model classifies objects into their correct categories. Initially, the class loss starts quite high, around 3.8, indicating poor classification accuracy in the early stages of training. However, as the epochs progress, the class loss rapidly decreases, particularly within the first 50 epochs, reflecting the model's significant improvements in learning to classify objects more accurately. After this steep decline, the loss continues to decrease at a slower rate, eventually stabilizing around 0.5 after approximately 100 epochs. This plateau suggests that the model has reached a point where further training does not significantly improve its classification accuracy. Although there are slight fluctuations in the later epochs, the overall trend shows that the model has achieved a high level of stability and robustness in classifying objects correctly, with minimal variation as training proceed.

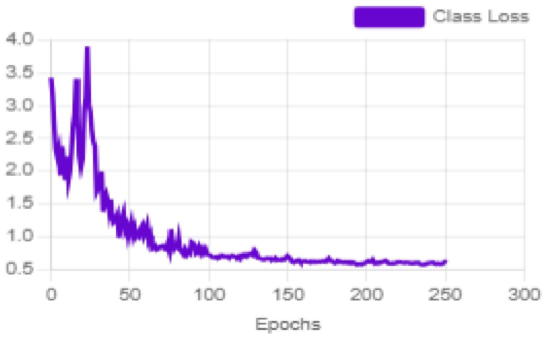

Fig. 10. Class Loss

A. *Object Loss*

The graph in Fig. 11 illustrates the 'Object Loss' over 300 training epochs, a metric in object detection that reflects the model's ability to detect and localize objects within an image. Initially, the object loss begins at a relatively high value of around 2.4, indicating that the model struggles with accurately detecting objects in the early stages of training. However, as the training progresses, there is a rapid decrease in object loss, especially within the first 50 epochs, showing that the model is quickly learning to improve its detection accuracy. After this sharp decline, the loss stabilizes between 1.7 and 1.9, starting around 100 epochs, with minor fluctuations. These fluctuations suggest that while the model's performance has largely converged, there may still be small inconsistencies in object detection. Nonetheless, the overall trend indicates that the model has achieved a relatively stable and robust detection capability, with the loss plateauing toward the end of the training process.

Fig. 11 Object Loss

VII. CONCLUSION

In conclusion, recent advances in deep learning and computer vision have shown substantial promise in applying motion detection systems within educational settings. The reviewed architectures including CNN, R-FCN, YOLO variants, and LSTMs have demonstrated effectiveness in tracking student behaviors, classifying motions, and even identifying emotions through real-time analysis. Such technologies enable more objective, efficient monitoring of student engagement and participation, offering a robust alternative to manual observation and traditional recording methods. Camera placement optimization, such as the use of overhead or multi-angle setups, further enhances detection accuracy and tracking precision, making these systems adaptable to complex classroom environments. Despite some challenges in precise localization and object detection stability at higher thresholds, the deep architectures discussed show reliable performance and convergence during training, indicating their readiness for broader implementation. The integration of AI-based motion detection holds the potential to transform educational monitoring, enabling instructors to foster more interactive and responsive learning environments tailored to students' needs.

ACKNOWLEDGMENT

"R. B. G. thanks Universiti Malaysia Kelantan (UMK), Kampus Bachok, and the Faculty of Science Data and Computing, Kampus Pengkalan Chepa, for providing the facilities and financial support for this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Chakraborty, P., Mittal, P., Gupta, M. S., Yadav, S., & Arora, A. (2020). Opinion of students on online education during the COVID-19 pandemic. *Human Behavior and Emerging Technologies, 3*(3), 357–365. https://doi.org/10.1002/hbe2.240.

[2] Cui, Y., Ma, Z., Wang, L., Yang, A., Liu, Q., Kong, S., & Wang, H. (2023). A survey on big data-enabled innovative online education systems during the COVID-19 pandemic. *Journal of Innovation & Knowledge, 8*(1), Article 100295. https://doi.org/10.1016/j.jik.2022.100295.

[3] Sahu, P. (2020). Closure of universities due to coronavirus disease 2019 (COVID-19): Impact on education and mental health of students and academic staff. *Cureus*. https://doi.org/10.7759/cureus.7541.

[4] Glazkova, G., Mamonova, O., Gracheva, D., & Pukhovskaya, M. (2020). Social aspects of physical education of students with deviations in health status. *Journal of Physical Education and Sport, 20*(5), 2545–2553.

[5] Li, Y., Qi, X., Saudagar, A. K. J., Badshah, A. M., Muhammad, K., & Liu, S. (2023). Student behavior recognition for interaction detection in the classroom environment. *Image and Vision Computing, 136*, Article 104726. https://doi.org/10.1016/j.imavis.2023.104726.

[6] Marwa, M., Nurfaisal, N., Mualiardi, M., Awal, R., & Irawan, H. (2024). Enhancing 21st century skills in the EFL classroom: Exploring teachers' implementation of group activities and assignments. *AL-ISHLAH: Jurnal Pendidikan, 16*(2), 1740–1756.

[7] Murtaza, M., Ahmed, Y., Shamsi, J. A., Sherwani, F., & Usman, M. (2022). AI-based personalized e-learning systems: Issues, challenges, and solutions. *IEEE Access, 10*, 81323–81342.

[8] Mao, M., & Chen, J. (2024). Research on the application technology of artificial intelligence in college physical education and training. *Applied Mathematics and Nonlinear Sciences, 9*(1).

[9] Qureshi, M. A., Riaz, A., Mughal, B., & Riaz, A. (2021). Optimizing camera placement for motion detection systems in educational environments. *Journal of Educational Technology Systems, 49*(3), 344–358.

[10] Cai, Y., Wang, Z., & Zhou, J. (2016). Robust human motion tracking with a camera network. *IEEE Transactions on Circuits and Systems for Video Technology, 26*(5), 865–879.

[11] Wang, J., Zhang, H., & Zhou, Z. (2021). Deep learning for real-time human action recognition in smart classrooms. *IEEE Transactions on Learning Technologies, 14*(2), 123–134.

[12] Zhou, J., Ran, F., Li, G., Peng, J., Li, K., & Wang, Z. (2022). Classroom learning status assessment based on deep learning. *Mathematical Problems in Engineering, 2022*, 1–9. https://doi.org/10.1155/2022/7049458.

[13] Saini, M. K., & Goel, N. (2019). How smart are smart classrooms? A review of smart classroom technologies. *ACM Computing Surveys, 52*(6), 1–28. https://doi.org/10.1145/3365757.

[14] Wang, Z., Yao, J., Zeng, C., Wu, W., Xu, H., & Yang, Y. (2023). Learning behavior recognition in smart classroom with multiple students based on YOLOv5. *arXiv*. https://arxiv.org/abs/2303.10916.

[15] Ramakrishnan, A., Zylich, B., Ottmar, E., LoCasale-Crouch, J., & Whitehill, J. (2021). Toward automated classroom observation: Multimodal machine learning to estimate class positive climate and negative climate. *IEEE Transactions on Affective Computing, 14*(1), 664–679.

[16] Carta, S. (Ed.). (2022). *Machine learning and the city: Applications in architecture and urban design*. John Wiley & Sons.

[17] Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., & Trautwein, U. (2021). Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology Review, 33*, 27–49.

[18] Zhou, J., Ran, F., Li, G., Peng, J., Li, K., & Wang, Z. (2022b). Classroom learning status assessment based on deep learning. *Mathematical Problems in Engineering, 2022*, 1–9. https://doi.org/10.1155/2022/704945.

[19] Hdioud, B., & Tirari, M. E. H. (2023). Facial expression recognition of masked faces using deep learning. *IAES International Journal of Artificial Intelligence, 12*(2), 921. https://doi.org/10.11591/ijai.v12.i2.pp921-930.

[20] Liu, M., Zhang, X., & Han, Y. (2020). Intelligent counting system for classroom numbers based on video surveillance. *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*. https://doi.org/10.1109/icpics50287.2020.9201999.

[21] Komagal, E., & Yogameena, B. (2023). PTZ-camera-based facial expression analysis using Faster R-CNN for student engagement recognition. In *Lecture Notes in Electrical Engineering* (pp. 1–14). https://doi.org/10.1007/978-981-19-7169-3_1.

[22] Xie, Y., Zhang, S., & Liu, Y. (2021). Abnormal behavior recognition in classroom pose estimation of college students based on spatiotemporal representation learning. *Traitement du Signal, 38*(1), 89–95. https://doi.org/10.18280/ts.380109.

[23] Li, L., Chen, C. P., Wang, L., Liang, K., & Bao, W. (2023b). Exploring artificial intelligence in smart education: Real-time classroom behavior analysis with embedded devices. *Sustainability, 15*(10), Article 7940. https://doi.org/10.3390/su15107940.

[24] Feng, W., Tang, J., & Liu, T. X. (2019). Understanding dropouts in MOOCs. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), 517–524. https://doi.org/10.1609/aaai.v33i01.330151.

[25] Kumar, G., Singh, A., & Sharma, A. (2023). Ensemble deep learning network model for dropout prediction in MOOCs. *International Journal of Electrical and Computer*

Engineering Systems, 14(2), 187–196. https://doi.org/10.32985/ijeces.14.2.8.

[26] Wang, H., Gao, C., Fu, H., Chen, Z.-H., Wang, Q., He, Z., & Li, M. (2023). Automated student classroom behaviors' perception and identification using motion sensors. *Bioengineering, 10*(2), Article 127. https://doi.org/10.3390/bioengineering10020127.

[27] Chen, H., & Guan, J. (2022). Teacher–student behavior recognition in classroom teaching based on improved YOLO-V4 and Internet of Things technology. *Electronics, 11*(23), Article 3998. https://doi.org/10.3390/electronics11233998.

[28] Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing, 7*(1), Article 48. https://doi.org/10.3390/bdcc7010048.

[29] Wang, Z., Yao, J., Zeng, C., Wu, W., Xu, H., & Yang, Y. (2022). YOLOv5 enhanced learning behavior recognition and analysis in smart classroom with multiple students. https://doi.org/10.1109/ieir56323.2022.10050042.

[30] Chen, H., Zhou, G., & Jiang, H. (2023). Student behavior detection in the classroom based on improved YOLOv8. *Sensors, 23*(20), Article 8385. https://doi.org/10.3390/s23208385.

[31] Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. In *Lecture Notes in Computer Science* (pp. 40–51). https://doi.org/10.1007/978-3-319-61425-0_4.

[32] Kumar, B. P. P. (2016). Performance analysis of KNN, SVM and ANN techniques for gesture recognition system. *Indian Journal of Science and Technology, 9*(1), 1–8. https://doi.org/10.17485/ijst/2017/v9is1/111145.

[33] Alsabhan, W. (2023). Student cheating detection in higher education by implementing machine learning and LSTM techniques. *Sensors, 23*(8), Article 4149. https://doi.org/10.3390/s23084149.

[34] Carrara, F., Elias, P., Sedmidubsky, J., & Zezula, P. (2019). LSTM-based real-time action detection and prediction in human motion streams. *Multimedia Tools and Applications, 78*(19), 27309–27331. https://doi.org/10.1007/s11042-019-07827-3.

[35] Xi, J. (2022). Research on flipped classroom of university curriculum using eye movement analysis and LSTM neural network. *Wireless Communications and Mobile Computing, 2022*, 1–10. https://doi.org/10.1155/2022/2559864.

[36] Li, X. (2021). Design and implementation of human motion recognition information processing system based on LSTM recurrent neural network algorithm. *Computational Intelligence and Neuroscience, 2021*, 1–9. https://doi.org/10.1155/2021/3669204.

[37] Qayyum, M., Nazeer, M., & Abdul Ahad. (2022). Real time object detection and recognition in machine learning using Jetson Nano. *ISSN 2456–5083*, 118. https://www.researchgate.net/publication/369595618.

[38] Yang, F., Ismail, N. A., Pang, Y. Y., Kebande, V. R., Al-Dhaqm, A., & Koh, T. W. (2024). A systematic literature review of deep learning approaches for sketch-based image retrieval: Datasets, metrics, and future directions. *IEEE Access, 12*, 14847–14869. https://doi.org/10.1109/access.2024.3357939.

[39] Sharma, V., Gupta, M., Pandey, A. K., Mishra, D., & Kumar, A. (2022). A review of deep learning-based human activity recognition on benchmark video datasets. *Applied Artificial Intelligence, 36*(1). https://doi.org/10.1080/08839514.2022.2093705.

[40] Sharma, V., Gupta, M., Kumar, A., & Mishra, D. (2021). Video processing using deep learning techniques: A systematic literature review. *IEEE Access, 9*, 139489–139507. https://doi.org/10.1109/access.2021.3118541