Sage Science Review of Educational Technology
I.2.10 Vision and Scene Understanding: Classification and scene analysis.
I.2.6 Learning: Learning algorithms and architectures.
K.3.1 Computer Uses in Education: Computers and education.
I.5.4 Applications: Applications of image processing and computer vision.

Student Engagement Detection in Classrooms through Computer Vision and Deep Learning: A Novel Approach Using YOLOv4

Aravind Sasidharan Pillai

Principal Data Architect, Data Engineering, Cox Automotive Inc., USA. ORCID: 0000-0001-7139-2804. Alumnus, Master in Data Science, University of Illinois Urbana-Champaign.

Abstract

Identifying early signs of disengagement in students is critical for reducing dropout rates in educational institutions. This study introduces a novel model that utilizes computer vision and deep learning to monitor student engagement levels within the classroom environment. The proposed model employs a camera system to capture images within the classroom at random intervals, which are then associated with student IDs and analyzed to assess engagement. This analysis categorizes the images into *'engaged'* and *'not engaged'*, from which a weekly average engagement score is derived and proposed to be communicated to each student. The differentiation between student images and general classroom activity is achieved using the YOLOv4 algorithm, trained on a custom dataset. YOLOv4, known for its balance between accuracy and processing speed, was selected as the primary tool for one-stage object detection. The dataset comprised 893 training images, 191 validation images, and 192 testing images. The study also explored a comparative analysis with a model trained on YOLOv3, examining performance metrics before and after augmentation with a Generative Adversarial Network (GAN)-enhanced dataset. Performance evaluation revealed improvements across all metrics after integrating GAN images, with YOLOv4 showing superior performance over YOLOv3. Notably, Mean Average Precision (mAP50) and Average Intersection over Union (IoU) saw significant gains, along with increases in Precision, Recall, and F1-score for both engaged and not engaged classifications. Privacy and security considerations are thoroughly addressed in the concluding section of the paper. The proposed model of this research shows the promise of these technologies in fostering a supportive learning environment by identifying and mitigating student disengagement.

Keywords:

computer vision, deep learning, student engagement, YOLOv4 algorithm, engagement monitoring, performance metrics, privacy considerations

Introduction

Kennelly and Monrad (2007) argue that students at risk of dropping out are typically disengaged in the classroom [1]. Therefore, they emphasize the importance of teachers observing and recognizing student behavior as a crucial measure to decrease dropout rates. Student engagement is a multifaceted concept that encompasses the investments of time, effort, and resources by both students and educational institutions. The goal is to enhance the overall student experience, improve learning outcomes, and contribute to the personal development of students. This synergy also boosts the institution's performance and reputation, aligning educational processes with broader academic objectives [2], [3].

At its core, engagement refers to the depth and quality of a student's involvement in the educational process. This includes their connection with school-related people, activities, goals, and values, as well as their sense of belonging to the educational environment. Effective engagement results in a meaningful and productive relationship between students and their academic settings, fostering a supportive atmosphere that promotes success [4].

Several factors influence student engagement in the classroom, with teacher interaction playing a pivotal role. Engaged teachers are enthusiastic, knowledgeable, and responsive, creating a learning environment that is both supportive and challenging. Their ability to connect with students, clarify complex concepts, and tailor teaching strategies to diverse learning needs significantly affects engagement levels [5], [6]. Classroom environment also significantly impacts student engagement. A well-organized, safe, and inclusive classroom promotes a sense of belonging and encourages active participation. Factors such as the physical arrangement of the classroom, access to resources, and the general atmosphere set by the school culture contribute to how comfortable and engaged students feel. Additionally, the use of technology and interactive tools can enhance learning experiences and maintain student interest and interaction.

Peer relationships and group dynamics are crucial as well. Students are more engaged when they feel part of a community of learners that supports mutual growth. Positive interactions among peers help build a collaborative spirit, reduce feelings of isolation, and encourage healthy competition. Establishing group norms and cooperative learning activities can foster this sense of community, enhancing engagement by making learning a more social and interconnected experience.

Student engagement embodies behavioral, emotional, and cognitive aspects, each playing a crucial role in a student’s educational experience. Behavioral engagement refers to the visible actions students take part in during both academic and non-academic activities. This includes interactions like asking questions, participating in class discussions, and adhering to classroom rules, as well as involvement in extracurricular activities beyond school hours. These behaviors are critical as they reflect the student's participation and conformity to the structured environment of educational settings, which directly impacts their learning outcomes and personal development.

On the emotional front, engagement is characterized by the students' affective reactions to the classroom setting and the instructional methods used. Emotional engagement is manifested through feelings such as enjoyment, anxiety, or boredom, which are influenced by the classroom dynamics and the pedagogical approaches employed. This type of engagement is vital because it affects how students relate to the content and their overall emotional well-being in the school environment. Meanwhile, cognitive engagement involves deeper mental processes, where students actively employ strategies like planning and self-monitoring to tackle academic challenges. This aspect of engagement not only enhances their ability to absorb and apply information but also fosters resilience and sustained motivation in the face of academic hurdles, thereby supporting their long-term educational success [7], [8].

Efforts to engage students effectively in the classroom primarily focus on managing behaviors to minimize disruptions and disciplinary issues, rather than solely addressing misconduct. A significant factor in enhancing student engagement is managing to capture and maintain disengaged students' interest through interactive and inclusive teaching methods. This approach shifts the focus from reactive discipline to proactive engagement, emphasizing the importance of creating a learning environment that fosters mutual respect and interest. Students’ perception of their teacher and the tasks at hand play pivotal roles in this dynamic. The relationship students perceive with their teacher—whether they feel cared for and supported—substantially influences their willingness to engage. Teachers who demonstrate positive behaviors and a genuine interest in student welfare typically inspire higher levels of engagement and positive responses from their students [9], [10].

The nature of the tasks assigned in class also critically impacts student engagement. Task choice should consider the difficulty level, the manner of instruction, and the availability of resources, which together can make learning either a stimulating challenge or a discouraging chore. Engaging tasks are typically those that incorporate problem-solving in groups, involve fun and interesting activities, and are relevant to the students' interests. These factors not only help in reducing classroom behavior problems but also enhance the quality of student engagement. By integrating interests into the curriculum and employing varied instructional strategies, such as small-group work and project-based learning, teachers can significantly improve students' enjoyment and active participation in their education, thereby fostering a more dynamic and effective learning environment.

Student engagement is influenced by a myriad of individual and environmental factors. Individual factors encompass personal characteristics unique to each student, such as emotional states, self-confidence, inherent motivation, and demographic specifics like minority group status. For instance, students from minority groups may experience lesser engagement due to societal pressures and potential marginalization, which can lead to higher dropout rates. Additionally, students with special needs often require tailored educational approaches and resources, which, if not provided, may hinder their ability to engage fully with the learning process. These individual aspects are pivotal in shaping how students interact with their educational environments and participate in learning activities [11], [12].

Environmental factors play a critical role in either fostering or impeding student engagement. Among these, the influence of peers—especially through supportive friendships—is crucial. Friends who provide emotional and academic support can enhance a student's ability to engage with school activities positively. Similarly, the family environment and parenting styles significantly affect a student's behavior and engagement in school. Supportive families that encourage education can bolster student motivation and active participation in school activities [13]. These factors highlight the importance of a nurturing support system both within and outside the educational setting.

Interaction with teachers also significantly impacts student engagement. Teachers who create a supportive and inclusive classroom atmosphere can motivate students to participate more actively. This is facilitated by employing varied and engaging teaching methods that cater to diverse learning needs. The overall school climate, including the physical environment and the relationships among students and between students and staff, similarly influences engagement [14]. A positive and conducive school climate that promotes healthy interpersonal relationships and effective classroom management enhances student involvement in learning activities.

School rules and regulations also contribute to shaping a conducive learning environment. When students are involved in the creation and understanding of school rules, they are more likely to appreciate their importance and adhere to them, which promotes a structured and disciplined environment conducive to learning. Such engagement with the rules not only helps maintain order but also instills a sense of responsibility and community among students, fostering greater overall engagement [15].

Student engagement detection

In educational research, the detection of learners' engagement is categorized into three main approaches: automatic, semi-automatic, and manual, each incorporating different levels of user involvement. Manual methods rely directly on the learners' input through self-reporting techniques or external observations [16], [17]. Self-reporting involves learners completing questionnaires that gauge their focus and emotional state, providing indirect indicators of engagement through descriptive latent variables. Observational checklists, another manual approach, use external observers to assess engagement based on behaviors and classroom activities. This method may include subjective opinions and objective measures, but it risks equating compliance with actual engagement [18], [19].

Table 1. **Approaches to Detecting Learners' Engagement in Educational Research**

Approach	**Description**	**Examples**
Manual Methods	Relies directly on learners' input through self-reporting techniques or external observations.	**Self-reporting**: Learners complete questionnaires gauging their focus and emotional state, providing indirect indicators of engagement. **Observational checklists**: External observers assess engagement based on behaviors and classroom activities, potentially including subjective opinions and objective measures.
Semi-automatic	Involves analyzing the timing and accuracy of responses to educational content, employing probabilistic inference to distinguish between engaged and disengaged states. Requires some indirect involvement from learners.	**Engagement tracing**: Analyzes response timing and accuracy to distinguish between engaged and disengaged states, predominantly used in intelligent tutoring systems.
Automatic Methods	Facilitates engagement detection without interrupting the learning process, utilizing data from computer vision, sensor analysis, and log files. Computer vision techniques parse physical cues like facial expressions and eye movements, while sensor data analyses focus on physiological and neurological signals indicating arousal levels associated with engagement. Log-file analysis provides insights from digital interactions within learning environments.	**Computer vision techniques**: Analyzing facial expressions and eye movements. - **Sensor data analysis**: Monitoring physiological and neurological signals indicating arousal levels. - Log-file analysis: Capturing detailed metrics like time allocation and activity rates.

Semi-automatic methods, such as engagement tracing, analyze the timing and accuracy of responses to educational content, employing probabilistic inference to distinguish between engaged and disengaged states [20]. These methods are predominantly used in intelligent tutoring systems and require some indirect involvement from learners. Automatic methods facilitate engagement detection without interrupting the learning process, utilizing data from computer vision, sensor analysis, and log files. Computer vision techniques parse physical cues like facial expressions and eye movements, while sensor data analyses focus on physiological and neurological signals indicating arousal levels associated with engagement [21], [22]. Log-file analysis provides insights from digital interactions within learning environments, capturing detailed metrics like time allocation and activity rates. These automatic approaches offer robust, non-invasive means to assess engagement, catering to the evolving dynamics of educational settings [23].

Computer vision-based methods analyzes cues such as gestures, postures, eye movements, and facial expressions. These methods are valued for their unobtrusive nature, allowing observations similar to how a teacher might assess motivation in a traditional classroom setting without disrupting student activities. The accessibility and affordability of camera technology, now commonplace in devices like cell phones, tablets, and computers, enhance the feasibility of using computer vision to detect engagement. Additionally, the integration of affective computing techniques further refines this approach. Despite the progress and potential of these technologies, particularly within the Intelligent Tutoring Systems (ITS) community, significant developments are still required to adapt these automated systems for broader educational applications, including online learning environments. This need for further innovation underscores the ongoing challenge of implementing computer vision comprehensively across diverse learning settings.

The objective of this research is to develop a novel model using computer vision and deep learning to identify early signs of disengagement in students, aiming to decrease dropout rates by monitoring and assessing student engagement levels in real-time within the classroom environment.

Method

Proposed system

The implementation of a computer vision and deep learning model for monitoring student engagement in the classroom involves several steps as described below and shown in diagram in Figure 1.

1. Camera System Setup- ***Camera Model***: The classroom is equipped with multiple high-resolution cameras, such as the Canon EOS 5D Mark IV. These cameras are chosen for their ability to capture detailed images in various lighting conditions. ***Placement***: Cameras are strategically placed around the classroom to cover all angles, ensuring that every student's face can be captured without obstructions. Ideally, at least one camera is positioned at the front, back, and sides of the room.

2. Image Capture Process- ***Random Interval Generation***: The system utilizes a software scheduler that generates random time intervals within a specified range for image capture. For example, it might be programmed to trigger the camera every 10 to 20 minutes during class hours. Capture ***Mechanism***: Upon reaching the randomly determined interval, the cameras simultaneously capture images of the classroom. This ensures a diverse dataset representing various moments within a class session.

3. Image Processing and Student Identification- ***Preprocessing***: Captured images undergo preprocessing to enhance quality and facilitate analysis. This includes steps like resizing, normalization, and lighting correction. ***Student Detection and Identification***: A pre-trained deep learning model, such as YOLO (You Only Look Once) or Faster R-CNN, is used to detect students' faces. Each face is then associated with a student ID using facial recognition technology, ensuring that the data is accurately tracked for each individual.

4. **Engagement Analysis-*****Engagement Assessment Model***: The model for assessing engagement could be based on a neural network architecture trained on a labeled dataset of 'engaged' and 'not engaged' student images. The network might use features such as eye gaze direction, facial expression, and body posture to assess engagement. **Categorization**: Each captured image is analyzed by the model, and the detected faces are categorized into 'engaged' or 'not engaged' based on the learned features.

5. **Calculation of Weekly Engagement Score-** ***Score Derivation***: For each student, the total number of 'engaged' categorizations is divided by the total number of images in which the student was detected over the week. This ratio represents the weekly engagement score. ***Random Check Interval***: To ensure accuracy, the system also includes a manual review process at random intervals, say once every two months, where a random sample of images is checked by human observers.

6. Feedback Mechanism-***Communication***: A secure, automated system communicates the weekly engagement scores to students via email or a dedicated educational platform. This feedback includes not only their score but also tips for improving engagement if necessary. ***Privacy Considerations***: The system is designed with strict privacy controls, ensuring that images and engagement scores are only accessible to authorized personnel (e.g., the individual student and designated educators).

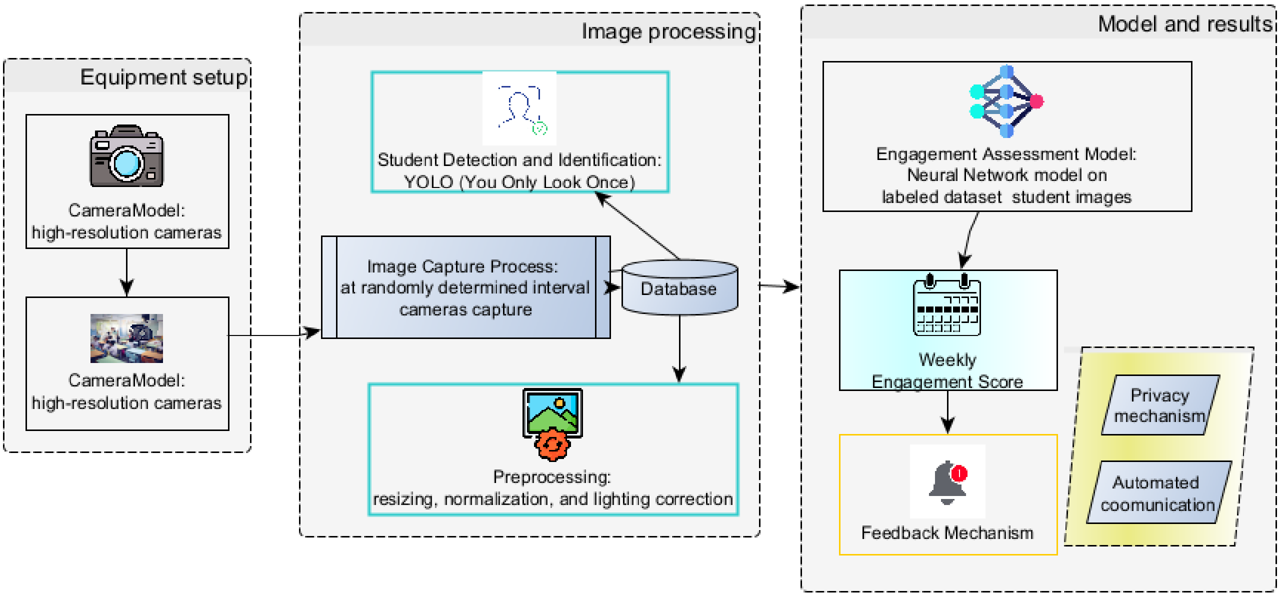

Figure 1. Architecture of the proposed system Source: Author

Computational methods

YOLO (You Only Look Once) represents a significant advancement in the field of object detection algorithms. Its unique approach frames object detection as a regression problem, consolidating the prediction of bounding boxes and class probabilities into a single neural network evaluation. This approach simplifies the detection process and reduces computation time compared to traditional methods, making it a notable development in the computer vision community. In the YOLO system, images are resized and processed through a single convolutional neural network, which predicts multiple bounding boxes and their associated class probabilities simultaneously. Post-processing techniques such as non-max suppression are then applied to refine the detections. While YOLO excels in terms of speed and efficiency, it may encounter challenges in precisely localizing small objects within images, a limitation inherent to its regression-based approach. Despite its limitations, YOLO's efficiency and real-time performance have made it a popular choice for various applications, ranging from surveillance systems to autonomous vehicles. Its impact on the field of object detection cannot be understated, as it has spurred further research and advancements in neural network architectures. However, like any algorithm, YOLO has its strengths and weaknesses, and its suitability depends on the specific requirements of the task at hand.

Object Detection Unified

Within the context of YOLO (You Only Look Once), input visuals are segmented into a grid of S × S dimensions (refer to Figure 9). Should an object's central point reside within a specific grid segment, that segment assumes the responsibility for object detection. Each segment within the grid forecasts bounding frameworks (B) along with confidence ratings for these frameworks through leveraging image-wide features. These confidence ratings signify both the likelihood of an object's presence within the framework and the precision of the object type forecast. The formula for calculating the confidence rating is as follows:

$$\text{confidence} = Pr(\text{class}|\text{object}) \times Pr(\text{Object}) \times \text{IoU}_{\text{truth}}^{\text{pred}}; \quad Pr(\text{object}) \in [0,1]$$

In this equation, $Pr(\text{object})$ denotes the probability that any given grid segment houses an object, while $Pr(\text{class}|\text{object})$ indicates the likelihood of a particular object's presence within a segment, assuming the segment indeed contains an object. A $(Pr(\text{object}))$ value of 0.5 suggests a 50% probability of the framework enclosing an object. The term $(\text{IoU}_{\text{truth}}^{\text{pred}})$ represents the Intersection over Union metric applied to actual versus predicted bounding frameworks. A confidence rating of zero implies an absence of objects within the grid segment. This confidence rating is integral to the mean Average Precision (mAP) calculation at a designated threshold, wherein bounding frameworks with confidence ratings beneath the set threshold are disregarded. Every bounding framework is characterized by attributes including $(bx), (by), width((w))$, height (h), and a confidence rating per object. The coordinates $(bx), (by)$ pinpoint the framework's center relative to the grid segment's boundaries, whereas the width $((w))$ and height $((h))$ are predicted in relation to the entire visual.

Modern object detectors typically consist of two main components: a backbone and a head. The backbone, pretrained on datasets like ImageNet, serves as the foundational feature extractor. Common backbones utilized on GPU platforms encompass VGG, ResNet, and DenseNet, while those tailored for CPU platforms include SqueezeNet, MobileNet, and ShuffleNet. These backbones extract essential features from input images. The head, on the other hand, is responsible for predicting classes and bounding boxes. Object detectors are further classified based on their heads into two types: one-stage and two-stage detectors. The former, exemplified by YOLO, SSD, and RetinaNet, directly predict object bounding boxes and class probabilities. In contrast, the latter, represented by R-CNN series, faster R-CNN, and R-FCN, employ a region proposal mechanism to detect objects before classifying them. Recent advancements include the integration of intermediary layers dubbed the "neck" between the backbone and the head. These layers, such as Feature Pyramid Network (FPN), Path Aggregation Network (PANet), BiFPN, and NAS-FPN, aggregate feature maps from multiple scales to enhance detection performance.

The evolution of object detection architectures has led to the refinement and diversification of detection models. By leveraging pretrained backbones and integrating sophisticated heads, modern detectors exhibit robustness and accuracy in object recognition tasks. The distinction between one-stage and two-stage detectors reflects varying approaches to object localization and classification. One-stage detectors prioritize efficiency and speed by directly predicting objects, while two-stage detectors focus on accuracy through a separate proposal generation step. The introduction of the neck component further enhances detection capabilities by facilitating multi-scale feature fusion. These advancements underscore the continual pursuit of improving object detection systems to meet the demands of diverse applications, ranging from autonomous driving to surveillance and beyond.

YOLOv4, a state-of-the-art object detection architecture, embodies a fully convolutional network design characterized by 110 convolutional layers [24]. Within this architecture, 66 convolutional layers are 1 × 1 in size, while 44 are 3 × 3. The input layer initiates with a 3 × 3 convolutional layer comprising 32 filters, processing images of dimensions 416 × 416 with three RGB channels. The output layer, a 1 × 1 convolutional layer with a stride and padding size of 1, encompasses 33 filters. YOLOv4's backbone, CSPDarknet53, adeptly extracts intricate features from input images, augmented by the Neck, which incorporates Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN) modules. These components effectively increase the receptive field and extract features across multiple scales, enhancing the model's object detection capabilities. The head of YOLOv3 is retained in YOLOv4 for object detection tasks [25]. Optimization is achieved through mini-batch gradient descent with momentum, ensuring efficient convergence during training. With an input size of 416 × 416 and 3 RGB channels, YOLOv4 boasts over 60 million parameters, facilitating intricate feature representation and object detection accuracy.

In addition to its architectural components, YOLOv4 integrates Bag of Freebies (BoF) and Bag of Specials (BoS) techniques to further enhance performance. BoF strategies applied to the backbone include Mosaic data augmentation and DropBlock regularization, augmenting the dataset and mitigating overfitting. BoS techniques for the backbone encompass Mish activation functions and Cross-stage partial connections (CSP), enhancing feature representation and information flow throughout the network. For the detector, BoF implementations comprise novel loss functions like CIoU-loss, alongside techniques such as Class-Agnostic Mean Average Precision (CmBN), Self-Adversarial Training, and Random training shapes, optimizing model robustness and generalization. BoS enhancements for the detector include Mish activations, SPP-blocks, PAN-blocks, and Distance-IoU based Non-Maximum Suppression (DIoU-NMS), further refining object detection precision and recall. By amalgamating cutting-edge architectural elements with sophisticated optimization strategies [26], YOLOv4 stands as a pinnacle in real-time object detection, pushing the boundaries of accuracy and efficiency in computer vision applications.

Based on the split ratios of 70% for training, 15% for validation, and 15% for testing, this study allocated 1,276 images as follows:

Training set: 893 images,

Validation set: 191 images, and

Testing set: 192 images

Results

The table 2 presents the performance metrics for object detection models YOLOv4 and YOLOv3, both before and after the introduction of GAN-generated images. For both models, improvements are observed across all metrics after the use of GAN images.

For YOLOv4, the mean Average Precision (mAP50) increased from 74.08% to 77.08%, indicating a more accurate detection of objects. Similarly, the Average Intersection over Union (IoU) saw an improvement from 64.81% to 67.81%, suggesting better alignment between the predicted and actual bounding boxes. Precision, which measures the accuracy of positive predictions, rose from 77.76% to 80.76%. Recall, indicating the model's ability to find all relevant instances, improved from 76.34% to 79.34%. Consequently, the F1-score, which balances precision and recall, increased from 77.04% to 80.04%.

For YOLOv3, similar trends are evident. The mAP50 improved from 67.98% to 70.98%, and the Average IoU went up from 62.74% to 65.74%. Precision saw a rise from 79.36% to 82.36%, and Recall from 74.72% to 77.72%. The F1-score also saw a boost from 76.97% to 79.97%. These improvements demonstrate that the integration of GAN-generated images effectively enhances the detection capabilities of both YOLOv4 and YOLOv3 models, making them more precise and reliable in identifying objects within images.

Table 2. mAP, Precision, Recall, and F1-score for both classes

Metrics	**YOLOv4 Before**	**YOLOv4 After GAN images**	**YOLOv3 Before**	**YOLOv3 After GAN images**
mAP50	74.08%	77.08%	67.98%	70.98%
Average IoU	64.81%	67.81%	62.74%	65.74%
Precision	77.76%	80.76%	79.36%	82.36%
Recall	76.34%	79.34%	74.72%	77.72%
F1-score	77.04%	80.04%	76.97%	79.97%

The table 3 provides the performance metrics for the object detection models YOLOv4 and YOLOv3 on two classes, "Not engaged" and "engaged," both before and after the introduction of GAN-generated images.

For the class "Not engaged," YOLOv4 shows a significant improvement after incorporating GAN images, with the performance metric rising from 88.03% to 91.03%. YOLOv3 also benefits from the addition of GAN images, as its performance increases from 85.01% to 88.01%. These results indicate that the use of GAN-generated images enhances the models' ability to accurately detect and classify non-engaged instances.

Similarly, for the class "engaged," both models exhibit improvements when utilizing GAN images. YOLOv4's performance increases from 85.57% to 88.57%, while YOLOv3's performance goes up from 83.82% to 86.82%. This improvement underscores the effectiveness of GAN images in better capturing the nuances of engagement, leading to more accurate detection and classification by the models. The integration of GAN-generated images has clearly benefited both YOLOv4 and YOLOv3 across the two classes, enhancing their precision and reliability in object detection tasks specific to engagement status. This suggests that using such enhanced imaging techniques can be a valuable strategy for improving the performance of machine learning models in practical applications. 3 random input and associated output image from the system are shown in figure 2, 3, and 4.

Table 3. Per class model performance

Class	**YOLOv4 Before**	**YOLOv4 After GAN images**	**YOLOv3 Before**	**YOLOv3 After GAN images**
Not engaged	88.03%	91.03%	85.01%	88.01%
engaged	85.57%	88.57%	83.82%	86.82%

Conclusion

Understanding behavior data assists schools in determining areas of success, areas that need additional support, and the action steps required to promote students' independence. Schools can establish a supporting learning environment for every student by recognizing and addressing their behaviors. This study presents a novel approach to monitoring student engagement in classrooms using YOLOv4-based computer vision and deep learning, detailing enhancements in detection accuracy and potential applications for reducing student disengagement.

The dataset used comprises 893 training images, 191 validation images, and 192 testing images. This relatively small dataset size for deep learning standards could limit the model's ability to generalize across diverse classroom settings. Moreover, the diversity in student demographics and classroom environments might not be adequately represented, which can affect the accuracy and applicability of the engagement detection model in different educational contexts.

The decision to use YOLOv4 based on its balance between accuracy and speed is rational; however, the comparative analysis with YOLOv3 and the impact of GAN-augmented datasets primarily focus on performance metrics like precision and recall without thorough investigation into the real-world applicability and robustness of these models in actual classroom settings. : Classroom dynamics, including lighting conditions, the camera's placement, and uncontrolled classroom activities, can affect the quality of the image data and the subsequent analysis of engagement. These factors can introduce variability in the data that may not be fully accounted for in the model, affecting its accuracy and reliability.

The implementation of this system requires high-resolution cameras and substantial computational resources, particularly for processing and analyzing video data in real-time. This requirement may limit the feasibility of deploying this technology in schools with limited technological infrastructure. The system proposes to communicate weekly engagement scores to students, which involves assumptions about the motivational impact of such feedback. The effectiveness of this feedback mechanism in actually improving student engagement remains unproven and requires further empirical study to validate these claims. The continuous monitoring of students via camera can raise significant ethical concerns, including student consent and the potential for misuse of images. There is also a risk of bias if the dataset does not proportionally represent all demographic groups, potentially leading to skewed engagement assessments based on facial recognition and behavior analysis algorithms. The model's reliance on visual cues such as facial expressions, eye movements, and body postures can introduce errors. Such parameters can be influenced by cultural differences in nonverbal communication or by individual variations in expressiveness, potentially resulting in misclassification of engagement.

References

[1] L. Kennelly and M. Monrad, "Approaches to Dropout Prevention: Heeding Early Warning Signs with Appropriate Interventions," *American Institutes for Research*, 2007.

[2] E. R. Kahu, "Framing student engagement in higher education," *Studies in Higher Education*, vol. 38, no. 5, pp. 758–773, Jun. 2013.

[3] H. Coates, "The value of student engagement for higher education quality assurance," *Quality in Higher Education*, vol. 11, no. 1, pp. 25–36, Jan. 2005.

[4] J. D. Finn and K. S. Zimmer, "Student Engagement: What Is It? Why Does It Matter?," in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds. Boston, MA: Springer US, 2012, pp. 97–131.

[5] J. Parsons and L. Taylor, "Improving student engagement," *Current issues in education*, 2011.

[6] R. M. Carini, G. D. Kuh, and S. P. Klein, "Student engagement and student learning: Testing the linkages," *Res. High. Educ.*, vol. 47, no. 1, pp. 1–32, Feb. 2006.

[7] S. Christenson, A. L. Reschly, and C. Wylie, *Handbook of Research on Student Engagement*. Springer New York, 2012.

[8] V. Trowler, "Student engagement literature review," *The higher education academy*, 2010.

[9] S. Bird and M. Latimer, "Examining models of departmental engagement for greater equity," *Equal. Divers. Incl. Int. J.*, vol. 38, no. 2, pp. 211–225, Mar. 2019.

[10] Nudeshima J., "Ethical issues in artificial intelligence and neuroscience," *Brain Nerve*, vol. 71, no. 7, pp. 715–722, Jul. 2019.

[11] J. Reeve, "A Self-determination Theory Perspective on Student Engagement," in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds. Boston, MA: Springer US, 2012, pp. 149–172.

[12] G. D. Kuh, "Assessing What Really Matters to Student Learning Inside The National Survey of Student Engagement," *Change: The Magazine of Higher Learning*, vol. 33, no. 3, pp. 10–17, May 2001.

[13] J. A. Fredricks and W. McColskey, "The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments," in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds. Boston, MA: Springer US, 2012, pp. 763–782.

[14] S. R. Harper, *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations*. London, England: Routledge, 2010.

[15] J. J. Appleton, S. L. Christenson, D. Kim, and A. L. Reschly, "Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument," *J. Sch. Psychol.*, vol. 44, no. 5, pp. 427–445, Oct. 2006.

[16] M. M. Handelsman, W. L. Briggs, N. Sullivan, and A. Towler, "A Measure of College Student Course Engagement," *J. Educ. Res.*, vol. 98, no. 3, pp. 184–192, Jan. 2005.

[17] M. Wen, D. Yang, and C. Rose, “Linguistic Reflections of Student Engagement in Massive Open Online Courses,” *ICWSM*, vol. 8, no. 1, pp. 525–534, May 2014.
[18] S.-F. Lam *et al.*, “Understanding and measuring student engagement in school: the results of an international study from 12 countries,” *Sch. Psychol. Q.*, vol. 29, no. 2, pp. 213–232, Jun. 2014.
[19] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions,” *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 01 Jan-March 2014.
[20] A. Psaltis, K. C. Apostolakis, K. Dimitropoulos, and P. Daras, “Multimodal Student Engagement Recognition in Prosocial Games,” *IEEE Trans. Comput. Intell. AI Games*, vol. 10, no. 3, pp. 292–303, Sep. 2018.
[21] C. R. Henrie, L. R. Halverson, and C. R. Graham, “Measuring student engagement in technology-mediated learning: A review,” *Comput. Educ.*, vol. 90, pp. 36–53, Dec. 2015.
[22] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, “Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores,” *Comput. Intell. Neurosci.*, vol. 2018, p. 6347186, Oct. 2018.
[23] A. K. Saxena, “Balancing Privacy, Personalization, and Human Rights in the Digital Age,” *Eigenpub Review of Science and Technology*, vol. 4, no. 1, pp. 24–37, 2020.
[24] Z. Jiang, L. Zhao, S. Li, and Y. Jia, “Real-time object detection method based on improved YOLOv4-tiny,” *arXiv [cs.CV]*, 09-Nov-2020.
[25] S. Fan *et al.*, “Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOV4 network,” *Comput. Electron. Agric.*, vol. 193, p. 106715, Feb. 2022.
[26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv [cs.CV]*, 23-Apr-2020.