Journal of Computer Science

Original Research Paper

Cheating Detection in Examinations Using Improved YOLOv8 with Attention Mechanism

[1]Yan Zuo, [1]Soo See Chai and [2]Kok Luong Goh

[1]*Faculty of Computer Science and Information Technology, University of Malaysia Sarawak Unimas, Kota Samarahan, Sarawak, Malaysia*
[2]*School of Science and Technology, International University College of Advanced Technology Sarawak Icats University College, Malaysia*

Article history
Received: 06-05-2024
Revised: 31-05-2024
Accepted: 05-08-2024

Corresponding Author:
Soo See Chai
Faculty of Computer Science and Information Technology, University of Malaysia Sarawak Unimas, Kota Samarahan, Sarawak, Malaysia
Email: sschai@unimas.my

Abstract: Examinations are among the most widely used and effective methods for assessing knowledge mastery, both domestically and internationally, and are extensively used in various talent-selection processes. Currently, offline exam venues usually rely on on-site manual invigilation combined with exam-monitoring videos to further strengthen invigilation efforts. However, this invigilation method not only utilizes large amounts of human and material costs but also cannot comprehensively detect cheating behavior during exam processes and thus fairness cannot be guaranteed. To improve the efficiency of video reviews during invigilation, save labor costs, and strengthen invigilation efforts, this study proposes the use of target detection algorithms to achieve automatic detection of cheating actions in the exam room. To improve the speed of video detection, a student's abnormal-behavior detection method was proposed based on improved YOLOv8 and attention mechanism to achieve real-time detection of cheating actions in an exam room on a regular performance computer. The results showed that the detection accuracy of the improved YOLOv8 model reached 82.71%, thus meeting the application requirements.

Keywords: Examinations, Student Abnormal Behavior, Detection, Improved YOLOv8

Introduction

In recent years, with the occurrence of global emergency, and instances of cheating in examination rooms and in traditional examinations, the detection of cheating is mostly done manually by human proctors or more lately through technology-supported solutions. A typical implementation method is the use of video surveillance systems (Kaddoura and Gumaei, 2022; Bergmans *et al.*, 2021) whereby proctors observe examinees via live streaming or recordings. Nevertheless, these systems continue to depend on human diligence that may be compromised due to sleepiness, distraction, or prejudice (Jackson, 1961). Some systems have made use of computerized technologies, such as the keystroke dynamics analysis (Pusara and Pu, 2004), whereby a unique key pattern is used to authenticate the test-taker. However, this method is limited to computer-based tests and does not specifically address the issue of cheating during a test. Other biometric systems based on facial and iris recognition are used (Awad, 2010), but they can sometimes fail because of poor lighting conditions or cause privacy concerns Automatic processing software has gained popularity significantly, especially with the rise of online tests. This type of software employs a machine-learning approach in determining student behaviors when cheating indicators are examined (Alsabhan, 2023; Abbas *et al.*, 2022). However, many such programs incorrectly identify false positives or negatives due to their inability to interpret comprehensively all human behaviors and the situations surrounding these behaviors (Resta and Laferrière, 2008). A number of studies have suggested employing machine-learning algorithms, such as Support Vector Machine (SVM) and neural networks to detect cheating (Genemo, 2022). While these methods can handle vast amounts of information and learn intricate patterns, they frequently have low interpretability and necessitate massive labeled data to operate effectively (Ribeiro *et al.*, 2016). With the application of object detection algorithms, including You Only Look Once (YOLO), there is great potential for cheat detection. YOLO is a sophisticated machine-learning algorithm that detects objects in images and video sequences (Redmon *et al*., 2016). Meanwhile, the first generations of YOLO were not capable of identifying small or ambiguous objects whose detection might be critical to determining subtle cheating behaviors (Khan *et al*., 2018). Accordingly, this study introduces a novel YOLOv8 algorithm based on an attention mechanism that can help overcome these drawbacks. By leveraging the attention mechanism, it is possible that a model can concentrate more on significant regions of an image or video, which may improve its performance in detecting minor cheating behaviors.

This study determined the current action labels based on video sequences to detect cheating during examinations. In addition, a comprehensive framework that can detect and classify anomalous behaviors and actions occurring within a test room is presented, potentially indicating instances of cheating. This framework was developed based on YOLOv8 and the attention mechanism. Owing to the unavailability of an open-source dataset specifically ensuring epidemics and natural disasters, has underscored that students have access to a cheat-free environment. Together with others, we aim to acquire the most recent technologies for primary schools. Designed for detecting cheating in paper-based tests, we manually built and compiled a dataset that was developed to demonstrate the various means by which students could deceive a proctor during a paper-based examination. The dataset comprises the most frequent cheating techniques, which are classified into six types: Normal, passing items (notes, rulers, etc.), whispering, putting hands under the table, taking out mobile phones, walking around, and looking around.

The main contributions of this study are as follows:

1. The paper introduces an improved version of the YOLOv8 algorithm integrated with an attention mechanism. This novel approach is specifically tailored to detect subtle cheating behaviors in examination settings by focusing on relevant image regions and enhancing the model's feature extraction capabilities
2. To address the absence of publicly available datasets for the detection of student cheating behaviors, the authors created a comprehensive custom dataset. This dataset captures a wide array of student actions during paper-based exams, including both normal behaviors and various cheating methods. It serves as a crucial resource for training and validating the proposed detection models
3. The paper demonstrates the effectiveness of the improved YOLOv8 model through extensive testing. The model achieved a high detection accuracy of 82.71%, proving its capability for real-time cheating detection in examination environments. This result underscores the practical applicability of the proposed method in enhancing examination security and integrity.

Related Work

One of the new approaches in the computer vision field is object detection (Wu *et al*., 2020) used for finding and localization of objects on images or video scenes. It does not only specify the present objects but also describes their bounding boxes, or rectangular space that includes targeted objects Fig. (1) shows object detection.

In this instance, the object detection algorithm detects an image with a bounding box that represents its position and size. As shown in Fig. (1) the algorithm detects an object associated with possible cheating behavior. Corresponding to real-time applications, such as video surveillance or autonomous driving, etc., accuracy and efficiency in object detection algorithms are highly associated.

The conventional algorithms in object-detection research have been replaced by deep learning models (Mittal *et al*., 2020). Those include a multi-feature fast pedestrian detection algorithm based on an adapted version of an oriented histogram gradient developed by Hong *et al*. (2016) using a discrete wavelet transform, as well as the scale-invariant key point and distinctive image feature presented in (Lowe *et al*., 2004). To achieve recognition, the application of SVMs and AdaBoost suggests that features should be computer-generated from information provided by hand to reflect various experimental situations. Additionally, these traditional object-detection algorithms have significantly more complicated feature extraction procedures compared to deep learning. The model performs a poor generalization as compared to the object identification techniques using deep learning (Hussein *et al*., 2022).

To begin with, the object detection problem was addressed using deep learning under a newer RCNN approach (Girshick *et al*., 2016). An improved form of the previous approach using SPPNet (He *et al*., 2015), provides great accuracy to object detection as Fast R-CNN. However, these two approaches rely on selective search algorithms that are computationally very cumbersome and require high memory resources to identify regions. The newly-developed Faster R-CNN method (Ren *et al*., 2017) which brought about improvements in candidate area selection efficiency and applied anchor boxes with various scales does not meet the real-time detection need as well as cannot overcome slow detecting speed. Moreover, the SSD is a single-shot multi-box detector and YOLO (Li *et al*., 2022; Qiao *et al*., 2019) based on regression detection of objects. Although these methods are much less accurate, their detection time is significantly increased.

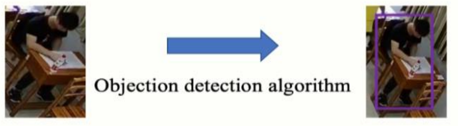

Fig. 1: Illustration of objection detection

DOI: 10.3844/jcssp.2024.1668.1680

A recent study has sought to enhance the above-described algorithm (Li *et al*., 2022). their paper incorporated a vehicle detection system based on R-FCN that can recognize objects with different sizes and dynamic backgrounds under various weather conditions to overcome these limitations Using depth-first search, dimension clustering module loss function as well slid window segmentation detections were adjusted so localization accuracy of the object along with the capability for detecting smaller objects was but these alterations led to the noticeable reduction in recall rate. Qiao *et al*. (2019) focused on improving the prediction accuracy of Faster R-detection Convolutional Neural Network (CNN) by refining their training dataset and came up with a two-channel network for feature extraction purposes. Nevertheless, the rate of detection did not change meaning that real-time identification was impossible.

In the above study, research focuses mainly on algorithm designs that are specifically designed for specific applications avoiding adaptation to monitoring deviant behavior during examination. Using dynamic threshold mining, (Lin *et al*., 2015; 2020) proposed a cheating-detection method based on the behavioral indicators of item exchange. They used the iterative threshold algorithm to determine a variable boundary for identifying differential images. The basis of the analysis was the results of segmentation which were employed to implement background replacement and detect cheating in the examination room with a subtraction algorithm from that. However, they did not include test results limiting the capacity of a method to detect abnormal item exchange patterns. Dai *et al*. (2012; 2019) presented an abnormality detection method in the exam room. Other concepts such as behavior coverage area and 3D examination-room attention have improved other aspects of the field. A latent SVM was employed for building a model but the resulting accuracy and speed did not exhibit conspicuous advantages. Besides, the use of this model for practical examinations poses a significant limitation on detection range.

Several studies have attempted to apply YOLO-based models to cheating detection in online examinations. Laurisa *et al*. (2022) proposed a method using YOLOv4 to detect cheating behavior by identifying unauthorized objects, such as mobile phones or books, in the video feed. Similarly, (Alkhalisy and Abd, 2022) utilized a YOLOv5-based model to detect multiple cheating indicators, including the presence of additional persons in the examination area. Despite the success of YOLO-based approaches in detecting cheating behaviors, several shortcomings require further investigation and improvement. One notable limitation is the high rate of false positives and negatives, which can result in the misclassification of cheating instances. Additionally, their dataset was very limited, with only 3 C4 classifications. Yulita *et al*. (2023) use MobileNetV2 architecture for cheating detection. Alsabhan (2023) created a deep learning model using LSTM layers with dropout and dense layers to identify exam cheating among students.

This study used the state-of-the-art YOLOv8 algorithm and an attention mechanism to create models and detect students' abnormal behavior to solve the aforementioned poor precision and low real-time processing capability. The YOLOv8 algorithm was slightly enhanced to detect abnormal behavior throughout the test.

The most recent and cutting-edge YOLO model, YOLOv8 (Jocher and Qiu, 2023), can be utilized in applications such as object identification, image categorization, and instance segmentation. Ultralytics (Jocher and Qiu, 2023), who produced the influential YOLOv5 model, developed the YOLOv8. Compared with YOLOv5, YOLOv8 showed several architectural updates and enhancements.

In summary, the literature has seen significant progress with deep learning models that have improved the precision and efficiency of object detection systems. These models have surpassed traditional algorithms, offering better feature extraction and more accurate real-time processing capabilities. There is a growing body of work that addresses the specific challenges of detecting cheating behaviors during exams. This includes the development of tailored algorithms and methods that can identify subtle cheating indicators and unauthorized objects within examination settings.

YOLOV8

YOLO is an object-detection algorithm that represents one of the computer vision techniques designed to recognize and locate objects inside images or video captures. More specifically, the object-detection algorithms are designed to identify objects that belong to different categories and locate them in an image using bounding boxes around these identified pixels. The modern YOLOv8 is a good solution for different tasks of object recognition and image segmentation being fast, accurate, and easy to implement. It can be implemented on a wide variety of hardware platforms from CPUs to GPUs and requires large data for training.

YOLOv8 utilizes the Darknet, which is similar to the Res Net and provides the basic framework of the algorithm. To overcome the issue of network degradation in deep convolutional neural networks and build deeper architectures, a residual block is designed to establish efficient parameter transfer between certain layers. When Darknet is used for object detection, the FC layer is removed and convolutional layers are employed. Figure (2) provides a detailed illustration of the YOLOv8 network architecture.

DOI: 10.3844/jcssp.2024.1668.1680

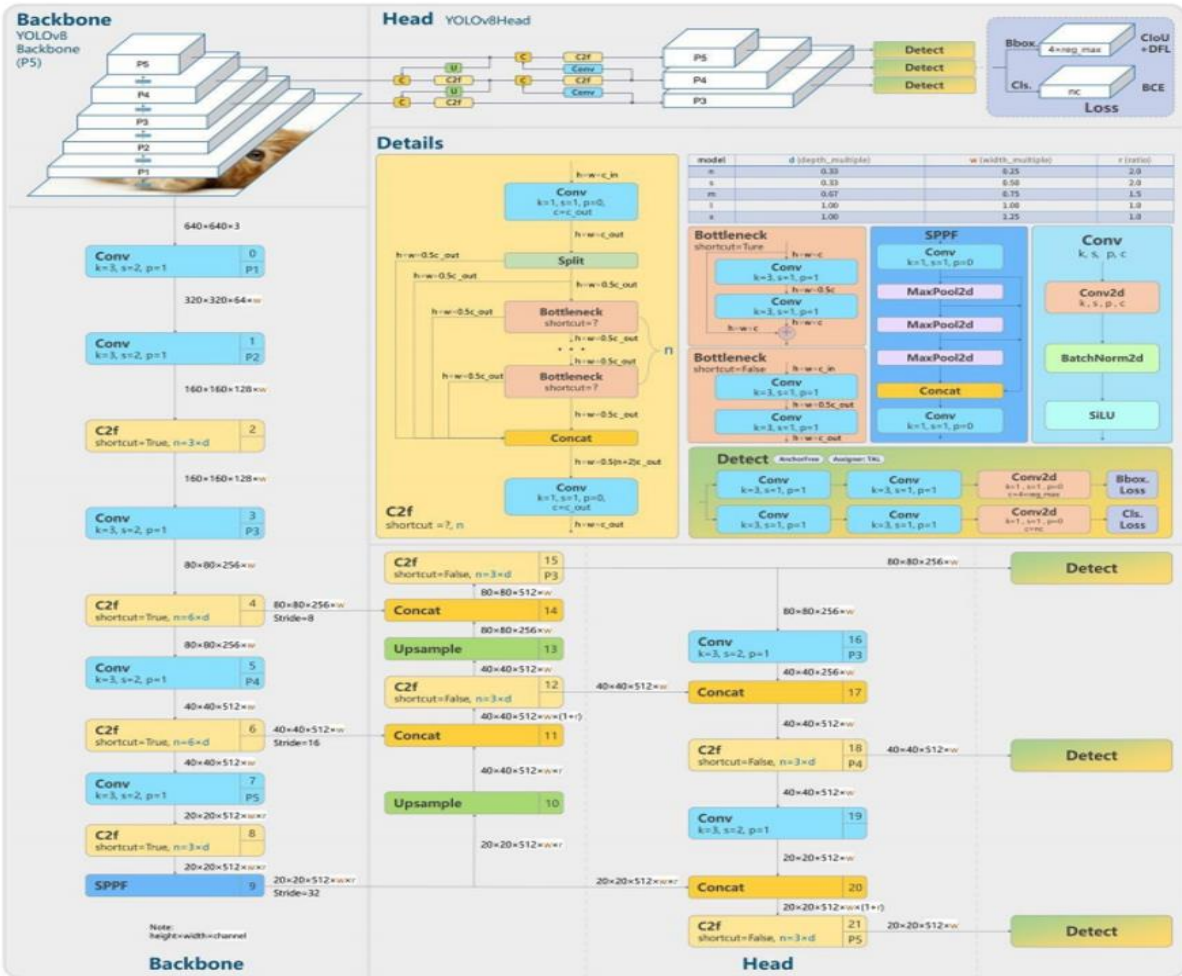

Fig. 2: YOLOv8 architecture

The YOLOv8 model was modified in several ways to improve its performance. Second, the C3 module was changed to a C2f module, and the initial 6×6 convolutional layer in the backbone was swapped with a 3 Moreover, the first 1×1 convolutional layer in the bottleneck was replaced with a 3×3 one. In addition, a severed head and abjectness branch were used. In particular, YOLOv8 is an anchor-free model that directly predicts the center of an object rather than estimating its offset from a predefined box. Figure (3) represents these modifications. The network predicts 4 coordinates for each bounding box, t_x, t_y, t_w, t_h. If the cell is offset from the top left corner of the image by (c_x, c_y) and the bounding box prior has width and height p_w, p_h, then the predictions correspond to.

In this study, anchor-free detection was utilized to reduce the number of box predictions, thereby speeding up the non-maximum suppression process, which is a complex post-processing step that involves sorting through candidate detections after inference Fig. (4).

In the final design of YOLOv8, the fundamental building block was modified by replacing C3 with C2f and the initial 6×6 convolutional layer in the stem was substituted with a 3×3 convolutional layer. Figure (5) illustrates the simplified network topology of YOLOv8, in which the backbone extracts the input image data and the head combines them to obtain more comprehensive target features for accurate predictions:

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{tw}$$
$$b_h = p_h e^{th}$$

Attention Mechanisms

In Computer Vision (CV), attention mechanisms are inspired by how humans selectively focus on specific parts of a scene while processing visual information. These mechanisms enable Neural Networks (NNs) to weigh the importance of different spatial regions or features within an image, helping the model focus on relevant areas and suppress irrelevant areas (Niu *et al*., 2021).

DOI: 10.3844/jcssp.2024.1668.1680

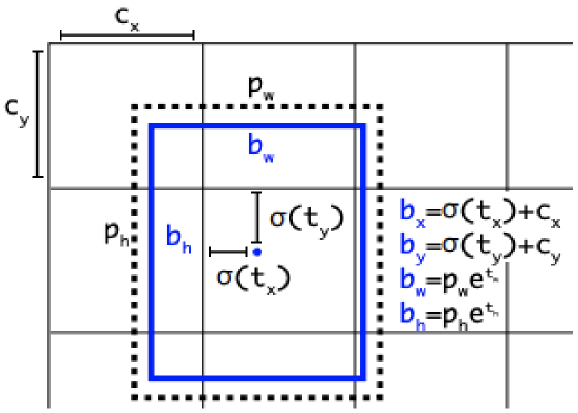

Fig. 3: Visualization of an anchor box in YOLO

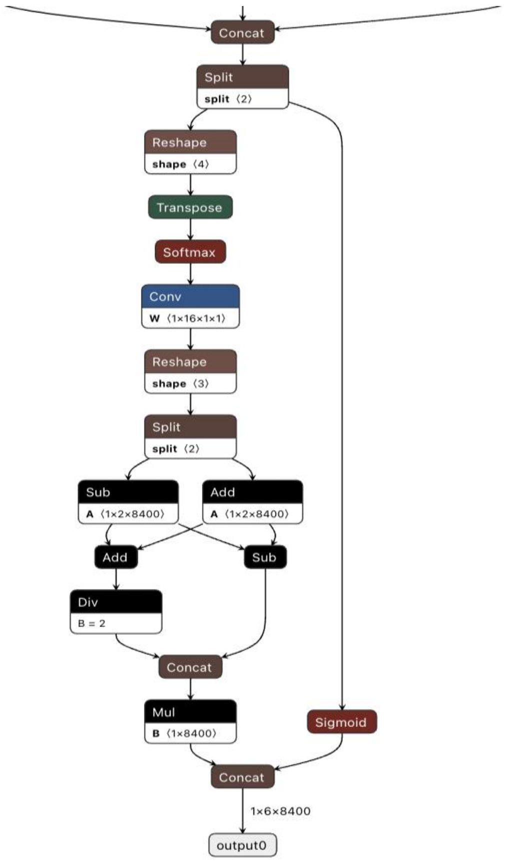

Fig. 4: Detection head for YOLOv8

Such mechanisms can be divided into three categories based on the scope of the action: Channel, spatial, and hybrid-domain attention. The channel attention mechanism involves Squeeze and-Exaction Networks (SENet). To generate the weights for studies collectively illustrate the transformative impact of attention mechanisms, heralding a shift in the paradigm of visual understanding and recognition.

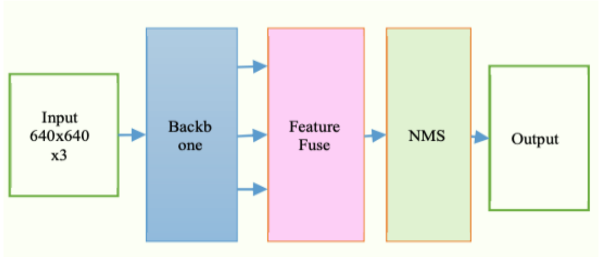

Fig. 5: Main flowchart of YOLOv8

The utilization of attention mechanisms in CV tasks has several advantages. First, attention mechanisms enhance the precision of object localization by focusing on relevant areas of an image. This is particularly critical in object detection, where accurate prediction of object boundaries is crucial. Additionally, attention mechanisms assist models in handling occlusions by guiding them to prioritize the nonoccluded portions of the image. This resilience to occlusions further enhances the performance and robustness of the models. For each channel, this mechanism employs an excitation operation comprising Fully Connected (FC) layers and a squeeze operation to aggregate feature information. In comparison, the Convolutional Block Attention Module (CBAM) and Bottleneck Attention Module (BAM) are hybrid attention mechanisms. To process the feature map, they combined the channel and spatial attention modules; however, BAM links the two modules in parallel and CBAM uses the channel-first priority to connect the modules in series.

Additionally, for feature aggregation, CBAM employs both max- and average pooling. Furthermore, in comparison to modules using a single dimension attention or single pooling method, CBAM can extract feature information that demands thorough attention. Figure (6) shows an overview of several typical attention methods.

Attention mechanisms have proven to be highly beneficial in image classification tasks by enabling the identification of crucial features within an image, thereby assisting the classification process. A prominent example is the Squeeze and Excitation (SE) network, which employs attention mechanisms to dynamically recalibrate feature responses on a channel-wise basis (Hu *et al.*, 2020). Attention mechanisms have been widely applied in object-detection tasks, particularly in models such as YOLOv3 and Focal Loss. These models utilize a feature pyramid network, which integrates a top-down attention mechanism. This attention mechanism allows the model to prioritize more salient regions of the image, resulting in improved object-detection accuracy (Wang *et al.*, 2022). In (Yu *et al.*, 2022), an SE network that explicitly modeled the interdependencies between channels in CNNs was proposed. Another study (Woo *et al.*, 2018) introduced CBAM to boost the representational power of the convolutional features. Furthermore, attention mechanisms are at the core of transformer architectures and have been successfully applied to various vision tasks.

DOI: 10.3844/jcssp.2024.1668.1680

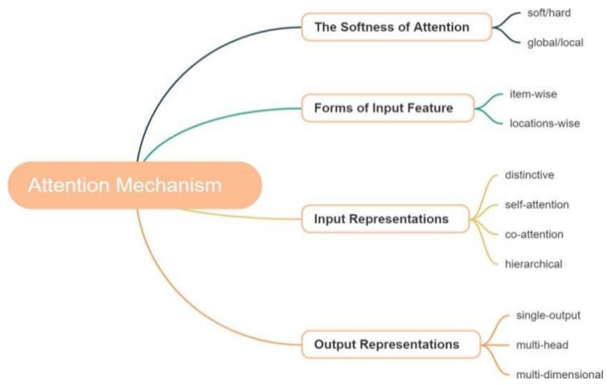

Fig. 6: Overview of several typical attention methods

Motivation

Despite its status as one of the most sophisticated variations on the YOLO architecture, it is not without some limitations. For instance, it shows a limited ability to capture the complex contextual arrangement present within images. This is due to the fact that YOLOv8 operates on a grid approach, in which each cell of frequency independently predicts bounding boxes and object classes. This type of approach may lead to the wasting contextual connections between different parts within an image. The incorporation of an attention mechanism successfully overcame these drawbacks. By explicitly and selectively modeling object-context relationships, attention mechanisms aid the model in identifying informative cues within an image that contribute to its ability to detect objects. This is accomplished by creating an attention map that brings to focus areas in the image that are considered important for effective detection and as a result minimizing the risk of missing key contextual information. By adding an attention mechanism to YOLO, the performance of this model can also be enhanced since it helps in focusing on more significant aspects of the input image. In addition, it has been demonstrated that attention mechanisms help to improve the performance of various deep-learning models mainly for CV tasks (Wang *et al*., 2022; Yu *et al*., 2022).

Meanwhile, despite the clear illustration of advantages found in combining attention mechanisms with object-detection models by the aforementioned studies, there is nothing less than an obvious gap in the application of YOLOv8 architecture. It surely lacks thorough research and experimentation integrating attentional modules within YOLOv8.

Cheating detection in an institutional context is a unique challenge. Conventional object detectors could sometimes miss subtle actions or objects related to cheating activities. On the other hand, attention mechanisms applied to YOLOv8 can make cheating detection much better. By formulating object-context relationships explicitly and concentrating on the desired spaces, attention-augmented YOLOv8 can accurately discriminate subtle signs of cheating. For instance, such activities as suspicious movements, secret devices, and the use of notes may be spotted more efficiently.

As such, the addition of attention mechanisms to YOLOv8 offers a viable approach for improving the reliability and precision of cheating detection. This integration can balance the weaknesses of YOLOv8 and use attention mechanisms to strengthen academic honesty.

Materials and Methods

Dataset Collection and Preparation

To address the lack of publicly available datasets focused on students' anomalous behavior particularly cheating during exams, we created and prepared our own dataset. This dataset includes all possible behaviors that students may undertake during paper-based examinations including actions with which they could cheat. It also includes most cheating methods and everyday behavior such as normal, passing items (notes or rulers), whispering with each other placing hands under the table taking out a mobile phone walk from different places to observe anything. Figure (9) illustrates the different states represented in this dataset. A Sony 30-170 mm lens was used on a Sony A7M4 camera to film the scenes that took place in one of China's high school classrooms with the enrolment population size being 26 students. The camera provided 30 frames per second at a resolution of 1920×1080.

The given frame rate is good enough for the detection of actions even minor movements without missing important details. The subject regions of the hand were captured by a top camera. Inherent challenges and complications are associated with the suggested dataset due to a close relationship between some activities as well as actions that do not depend only on body movements. For instance, actions, such as "drinking water" or "resting on the desk", should be taken as normal activities, which demand more background messages in order to understand them from abnormal actions. Consequently, when conducting the dataset analysis one cannot only define these subject's body movements but also determine their approach to objects while revealing hidden intentions. Integration of these parameters leads to a better understanding of why the subject behaved in such a way and confirms its accuracy. # detection and classification of anomalous actions.

We generated two 20-min videos to simulate the examination, totaling 40 min and 72,000 images. The labeling of all images requires more than a million annotations, a considerable amount of time, and human resources. However, not all frames are equally important for action recognition and only a few key frames are essential for describing each category. From the first video, we randomly selected 500 frames and performed meticulous labeling by assigning seven distinct label types. The distribution of labels for each of the seven categories is as follows: 2593, 1126, 1456, 1124, 523, 556 and 569, as shown in Table (1).

DOI: 10.3844/jcssp.2024.1668.1680

Table 1: Numbers of each label

Label	Number train and validation (video 1)	Testing (video 2)
Normal	2593	1385
Passing items	1126	0673
Whispering	1456	0697
Placing hands under The table	1124	0539
Taking out mobile phones	0523	0264
Walking around	0556	0295
Looking around	0569	0261

For training and validation, we utilized 80% of the labeled data images as the training set, whereas the remaining 20% served as the validation set for each category/state. The second video was reserved as the test set with the corresponding number of labels, as indicated in Table (1). We evaluated the performance using the errors of the training and validation sets and the prediction accuracy of the validation set as our evaluation indicators.

Proposed Approach

The weights given by the attention mechanism reflect how relevant different sections of input data are to a specific task. Weights are assigned to the pixels or regions in CV. Although YOLOv8 is considered one of the best models for object detection, its cheating signs-detection performance could be improved by unifying the selected attention mechanism with the model structure (mainly via the feature extraction phase or backbone). Through extensive research and a thorough review of the literature, five well-established attention mechanisms were identified as most widely used in practical applications: SENets (Woo *et al*., 2018), CBAM (Wang *et al*., 2020), Efficient Channel Attention (ECA) (Hou *et al*., 2021), Coordinate Attention (CA) and Second-Order attention network (SOCA).

To keep it concise, we represent the ECA mechanism below. This mechanism addresses the shortcomings of existing channel attention methods through complex operations and high computational costs. The ECA mechanism is inspired by the notion of local cross-channel interaction and it reflects that spatially adjacent channels have semantically close information. The local interactions in this case are captured efficiently by the ECA mechanism through a one-dimensional convolutional network that greatly improves its computational complexity over traditional methods of channel attention. The ECA module uses a kernel size k that is adaptively calculated depending on the number of channels within an input feature map. This flexibility ensures that local dependencies in varying channel dimensions are appropriately captured with the compact model size. Figure (7) depicts the architecture of the ECA module.

When we initially inserted the attention mechanism module into the backbone network, we encountered a challenge: The initial weights of the backbone network were disrupted, resulting in a decline in the prediction performance of the network. To address this issue and stage, we can maintain the network's functionality and prevent any negative impact on its prediction capabilities. We detail the tests conducted and compare the performance of all the aforementioned attention mechanisms to select the most effective improvement for YOLOv8. The selected attention model was seamlessly integrated into YOLOv8, as outlined in the flowchart in Fig. (8). In this study, we sequentially integrate five distinct attention modules into the YOLOv8 backbone to enhance feature expressiveness and detection performance. The core of our proposed methodology involves the sequential implementation of multiple attention mechanisms on feature maps derived from the YOLOv8 backbone.

Then, we introduced a novel architecture that integrates multiple attention mechanisms, namely SE, CBAM, ECA, CA, and SOCA, with the YOLOv8 object detection framework.After refining the feature maps through a series of attention modules, these features are processed through the YOLOv8 head. The YOLOv8 head typically contains convolutional layers to produce final outputs corresponding to objectness scores, bounding box coordinates, and class predictions.

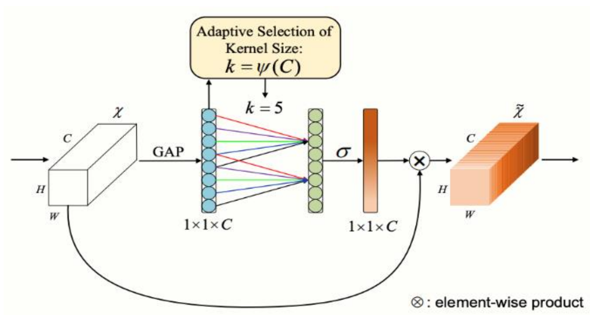

Fig. 7: Diagram of the ECA module. Given the aggregated features obtained by Global

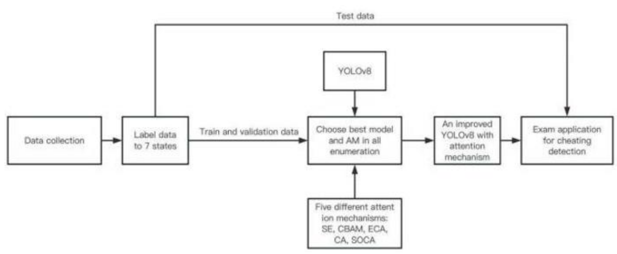

Fig. 8: Methodology of the study

Fig. 9: Seven different states: 1, Normal; 2, passing items (notes, rulers, etc.); 3, whispering; 4, placing hands under the Table (5), taking out mobile phones; 6, walking around; and 7, looking around captured in the dataset

The YOLOv8 loss function comprises three main components: Localization Loss: Measures errors in the predicted bounding box coordinates. Objectness Loss: Measures errors in the predicted objectness scores. Classification Loss: Measures errors in the class probability predictions.

After calculating the loss, we perform backpropagation to compute gradients with respect to the loss. These gradients indicate how much each parameter (weight) in the model contributed to the error.

Using the computed gradients, we then update the model's weights. This is typically done using optimization algorithms Adam. At the end of each training epoch, it is a common practice to validate the current model's performance on a separate dataset, ensuring that our model generalizes well to unseen data.

Results and Discussion

Experimental Setup

The experiment was conducted using a Windows 11 operating system equipped with an Intel (R) Core (TM) i7-8700 CPU @ 3.20 GHz with 32 cores, 128 GB of memory, and 2 NVIDIA GeForce GTX 3080 Ti GPUs.

The data preprocessing phase begins by labeling the images using the open-source software, LabelMe. This software enables the conversion of objects into JSON files containing coordinate information. To ensure compatibility of the labeled data with YOLOv8, a script was employed to convert the JSON files into a TXT format that can be read and utilized for training. Five models of YOLOv8 were launched, as demonstrated in Table (2). These models provide the following key information:

- Size (pixels): This parameter signifies the input resolution of the model

Table 2: Characteristics of the five YOLOv8 models

Model	size	mAP^{val}	Speed	Speed	Params	FLOPs
YOLOv8n	640	37.3	080.4	0.99	03.2	008.7
YOLOv8s	640	44.9	128.4	1.20	11.2	028.6
YOLOv8m	640	50.2	234.7	1.83	25.9	078.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

- mAPval 50-95: Mean Average Precision (mAP) is a standard metric for evaluating object detectors, such as YOLO. The "50-95" indicates that the map is computed over a range of intersection-over-union thresholds from 0.5-0.95, with a step size of 0.05
- Speed CPU ONNX (ms): This metric measures the inference speed of the model on a CPU using the Open Neural Network Xchange (ONNX) format
- Speed A100 Tensorrt (ms): As in the previous metric, this measures the inference speed of the model on an A100 GPU using NVIDIA's TensorRT framework. Lower values are preferable as they represent faster performance
- Params (M): This value represents the number of trainable parameters in the model measured in millions. Fewer parameters can be advantageous because they reduce overfitting and accelerate training and inference times. However, an excessively low number of parameters may limit the ability of the model to handle complex tasks
- FLOPs (B): Floating-point operations per second are a measure of the computational complexity. This indicates the number of operations the model must execute to make a prediction and is measured in billions. # These metrics provide crucial insights into the performance, efficiency, and computational demands of the YOLOv8 models. # As the dataset utilized in this study is unique, a comprehensive debugging and training process was conducted to evaluate all five attention mechanisms using the five YOLOv8 models. The objective is to identify the optimal model for a given dataset. Additionally, an additional set of five YOLOv8 models without attention mechanisms was trained for comparison and analysis

Data augmentation plays a vital role in enhancing the robustness of a model by enriching its training data with a diverse range of examples. This is particularly crucial in object-detection tasks, wherein objects can exhibit various shapes, sizes, orientations, and positions within an image. The default parameter settings were maintained for the data augmentation methods, including random cropping, random resizing and scaling, random flipping, random rotation, random brightness and contrast adjustments, and random hue and saturation adjustments. Additionally, the parameter settings of the YOLOv8 models were preserved at their default values.

The training process was conducted over 1000 iterations.

DOI: 10.3844/jcssp.2024.1668.1680

Results

The experimental results are presented in Table (3) revealing that the YOLOv8l model integrated with the ECA attention mechanism achieved the best performance. It exhibited the lowest training and validation loss at 933 epoch, as depicted in Fig. (10). However, as shown in the table, certain attention mechanisms employed with the YOLOv8 model are not as effective as those of the original YOLOv8 model, for example, the errors of the YOLOv8m model are always smaller than those of YOLOv8m-SE.

Figure (11) displays the confusion matrix based on YOLOv8l-ECA and the accuracy of various subclasses. Notably, even the subclass with the smallest accuracy, "taking out mobile phones," achieved an accuracy of 84.13%. The map demonstrates that the improved YOLOv8 model achieved a remarkable detection accuracy of 85.67%. These findings conclusively illustrate the efficiency of the training model and the effectiveness of the proposed method.

Consequently, the trained YOLOv8 model integrated with the ECA attention mechanism was employed to test the data in the second video. In general CV tasks, two evaluation metrics are commonly used, namely detection accuracy and detection speed. However, for this specific study, the detection accuracy takes precedence over the detection speed. The algorithm should be capable of detecting as many targets as possible, while maintaining an acceptable detection speed, such as achieving a smooth browsing experience with a minimum of 24 FPS. The detection speed is closely linked to the computer configuration and enhancing the capabilities of the computer can improve the overall detection speed. Figure (12) shows that the majority of the people in the video were detected with an FPS of 27. The improved YOLOv8 model achieved a detection accuracy of 82.71% for the labeled test set, that is, the confusion matrix Fig. (13), further emphasizing the efficiency of the training model and the effectiveness of the proposed method. Table (4) shows the detection accuracies of the classical objection-detection algorithms, SSD and Faster R-CNN, under the same configuration with the same test set. As shown, the improved YOLOv8 algorithm was superior to the other algorithms in Table (2). Characteristics of the five YOLOv8 models.

Table 3: Training results of YOLOV8 with an attention mechanism

Item	Train box_loss	Train obj_loss	Train cls_loss	Val box_loss	Val obj_loss	Val cls_loss
YOLOv8n	0.01933	0.00888	0.00496	0.00658	0.01816	0.01356
YOLOv8n-SE	0.02055	0.01185	0.01997	0.01553	0.01679	0.00962
YOLOv8n-CBAM	0.01840	0.01757	0.00282	0.01498	0.01085	0.0072
YOLOv8n-ECA	0.00211	0.01939	0.00758	0.01685	0.00794	0.01906
YOLOv8n-CA	0.00164	0.01968	0.01712	0.00772	0.00590	0.00981
YOLOv8n-SOCA	0.02054	0.00854	0.01397	0.00410	0.01866	0.01450
YOLOv8s	0.01545	0.01848	0.00210	0.01182	0.01018	0.01053
YOLOv8s-SE	0.01094	0.00772	0.02076	0.01435	0.00409	0.01392
YOLOv8s-CBAM	0.01588	0.01797	0.01976	0.01064	0.02041	0.00766
YOLOv8s-ECA	0.01799	0.01699	0.01101	0.00494	0.00958	0.01792
YOLOv8s-CA	0.00585	0.01205	0.01512	0.01604	0.01585	0.00522
YOLOv8s-SOCA	0.01032	0.0201	0.00909	0.00358	0.01008	0.01781
YOLOv8m	0.00712	0.00259	0.00140	0.00478	0.00579	0.00371
YOLOv8m-SE	0.01415	0.00845	0.00719	0.01875	0.00642	0.01673
YOLOv8m-CBAM	0.00647	0.01874	0.01282	0.00812	0.00233	0.00629
YOLOv8m-ECA	0.01750	0.00387	0.00577	0.00118	0.00540	0.01086
YOLOv8m-CA	0.02005	0.00832	0.00243	0.01817	0.01296	0.02069
YOLOv8m-SOCA	0.00797	0.00188	0.0020	0.00480	0.00634	0.00158
YOLOv8l	0.00252	0.01821	0.01138	0.00679	0.00916	0.01751
YOLOv8l-SE	0.01003	0.01527	0.01464	0.00232	0.01832	0.01756
YOLOv8l-CBAM	0.01766	0.01908	0.01689	0.00091	0.01011	0.00518
YOLOv8l-ECA	0.00101	0.00177	0.00135	0.00081	0.00097	0.00131
YOLOv8l-CA	0.00213	0.00241	0.01828	0.00669	0.02018	0.01772
YOLOv8l-SOCA	0.00915	0.00573	0.01823	0.01274	0.0068	0.01154
YOLOv8x	0.00763	0.01487	0.00471	0.01965	0.00623	0.00741
YOLOv8x-SE	0.00111	0.00187	0.00608	0.01743	0.00109	0.00357
YOLOv8x-CBAM	0.01777	0.00779	0.01708	0.01307	0.01626	0.01711
YOLOv8x-ECA	0.01952	0.01650	0.01223	0.02057	0.01928	0.01231
YOLOv8x-CA	0.01950	0.01815	0.01507	0.00135	0.01075	0.01369
YOLOv8x-SOCA	0.00955	0.00531	0.01522	0.01742	0.01342	0.00981

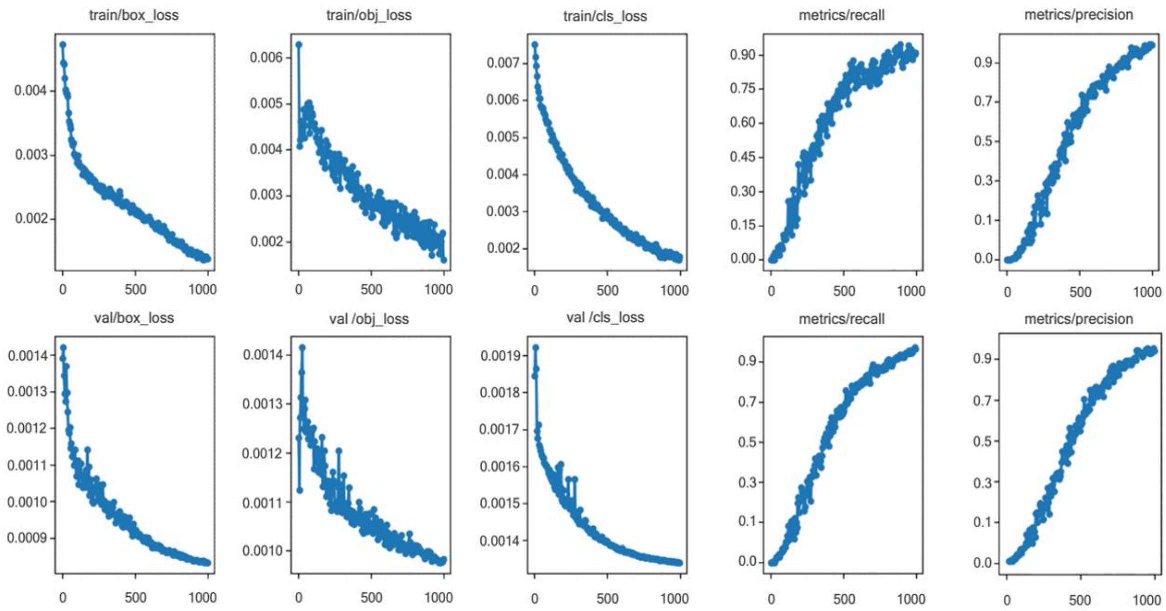

Fig. 10: Training results of YOLOv8l-ECA

Fig. 11: Traning results of YOLOv8l-ECA

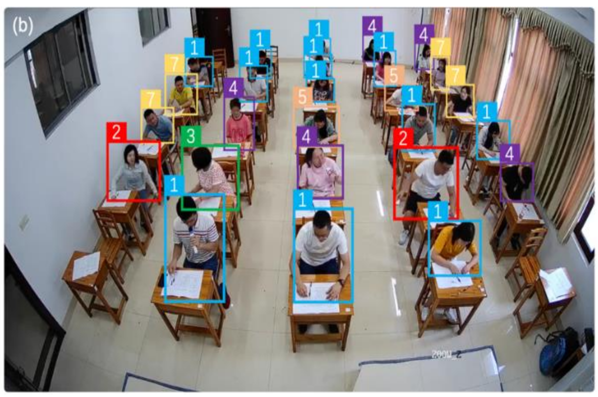

Fig. 12: Test results of YOLOv8l-ECA for two-event detection (a, b) [1: normal, 2: passing items (notes, rulers, etc.), 3: whispering, 4: placing hands under the table, 5: taking out mobile phones, 6: walking around, and 7: looking around]

Table 4: Comparison of different target-detection methods

Type	The improved YOLOv8	SSD	Faster R- CNN
Accuracy	82.71%	74.15%	73.39%
FPS	27	23	26

In specific applications, it is determined whether a student is engaging in cheating behavior based on the given context and circumstances. During testing, a significant majority of labels were accurately identified, with only a minimal number of instances, where labels were either missing or misidentified.

DOI: 10.3844/jcssp.2024.1668.1680

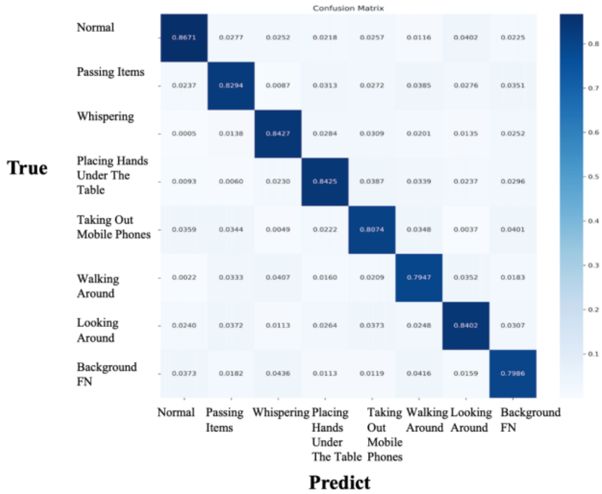

Fig. 13: Test results of YOLOv8l-ECA

Conclusion

This study addressed the challenges associated with detecting anomalous behavior in online examination centers by incorporating YOLOv8 with an attention mechanism. Specifically, the effectiveness of five attention mechanisms and YOLOv8 models was analyzed to optimize the algorithm. To facilitate this optimization process, we generated datasets specifically designed to analyze anomalous behavior during examinations. In particular, the proposed technique utilizes an enhanced YOLOv8l model with an ECA mechanism to process examination videos, enabling the evaluation of detection speed and accuracy. The results demonstrate that the combination of video surveillance in the exam room with the YOLOv8l model incorporating ECA enables real-time automatic cheating detection. The improved YOLOv8 algorithm for detecting abnormal behavior in examinations can provide valuable insights into the development of automated cheating-prevention systems. With emphasis on detection accuracy, the method achieved accuracy rates exceeding 85% in each subclass, further enhancing its effectiveness in detecting anomalous behavior.

The potential applications of the proposed approach are numerous. Primarily, it can be applied in educational settings for remote proctoring during online exams to ensure academic integrity in an increasingly digital world. By applying this approach, educational institutions can enhance fairness in examinations and mitigate potential misconduct. Moreover, this approach can have applications beyond the educational domain. For instance, it can be used in corporate settings to ensure adherence to ethical guidelines during online certification or training. Additionally, it may be used in the security industry, specifically for surveillance, to detect suspicious behaviors indicative of rule-breaking or illegal activity.

Note that the entire examination process is complex and not all unusual behaviors necessarily indicate cheating. Therefore, this process requires manual judgment. The method presented in this study has certain limitations, indicating room for further improvement. Future research endeavors will incorporate multitarget tracking technology, enabling precise behavior prediction for individuals rather than specific action types. By detecting the type and frequency of abnormal behavior exhibited by individuals throughout the examination, a more accurate and effective determination of cheating behavior can be achieved. Furthermore, the refinement of the attention mechanism holds significant potential for enhancing the model performance. Specifically, future investigations could explore the integration of transformer-based models, such as vision transformers, which utilize self-attention mechanisms to weigh the importance of different parts of the image. The inclusion of additional sensory data, such as audio or thermal imaging, could contribute to the development of a more comprehensive cheating detection system.

Acknowledgment

The authors express their sincere gratitude to all who contributed to the development of Cheating Detection in Examinations using Improved YOLOv8 with Attention Mechanism. Special thanks to our advisors, collaborators, and mentors for their invaluable guidance. This study is a testament to our collective effort and dedication. The authors also wish to thank Universiti Malaysia Sarawak for the financial support of this project.

Funding Information

The authors acknowledge Universiti Malaysia Sarawak (UNIMAS) through IMPACT Research Grant Scheme (F08/PARTNERS/2119/2021) for supporting this project.

Author’s Contributions

All authors equally contributed to this study.

Ethics

This manuscript is an original work. The authors declare that there are no ethical concerns associated with this submission.

Conflict of Interest

The authors have no competing interests to declare relevant to this article's content.

References

Abbas, M. A. E., & Hameed, S. (2022). A Systematic Review of Deep Learning Based Online Exam Proctoring Systems for Abnormal Student Behaviour Detection. *International Journal of Scientific Research in Science, Engineering and Technology*, *9*(4), 192–209. https://doi.org/10.32628/ijsrset229428

Alkhalisy, M. A., & Abd, S. H. (2022). Student Abnormal Behavior Detection Using Dlib Combined with YOLO Models. 2022 *International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IoE)*, 1–7. https://doi.org/10.1109/itss-ioe56359.2022.9990929

Alsabhan, W. (2023). Student Cheating Detection in Higher Education by Implementing Machine Learning and LSTM Techniques. *Sensors*, *23*(8), 4149. https://doi.org/10.3390/s23084149

Awad, A. I. (2010). An Efficient Iris Recognition System Based on Structural and Statistical Features. *Soft Computing*, *14*, 869–879.

Bergmans, L., Bouali, N., Luttikhuis, M., & Rensink, A. (2021). On the Efficacy of Online Proctoring using Proctorio. *Proceedings of the 13th International Conference on Computer Supported Education*, 279–290. https://doi.org/10.5220/0010399602790290

Dai, J. B., Long, M. L., Zhao, H. W., & Chen, F. J. (2012). Algorithm of the Exam Abnormal Behavior Detection. In *J. Jilin Univ*.

Dai, T., Cai, J., Zhang, Y., Xia, S.-T., & Zhang, L. (2019). Second-Order Attention Network for Single Image Super-Resolution. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11057–11066. https://doi.org/10.1109/cvpr.2019.01132

Genemo, M. D. (2022). Suspicious Activity Recognition for Monitoring Cheating in Exams. *Proceedings of the Indian National Science Academy*, *88*(1), 1–10. https://doi.org/10.1007/s43538-022-00069-2

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(1), 142–158. https://doi.org/10.1109/tpami.2015.2437384

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904–1916. https://doi.org/10.1109/tpami.2015.2389824

Hong, G.-S., Kim, B.-G., Hwang, Y.-S., & Kwon, K.-K. (2016). Fast Multi-Feature Pedestrian Detection Algorithm Based on Histogram of Oriented Gradient using Discrete Wavelet Transform. *Multimedia Tools and Applications*, *75*(23), 15229–15245. https://doi.org/10.1007/s11042-015-2455-2

Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate Attention for Efficient Mobile Network Design. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13708–13717. https://doi.org/10.1109/cvpr46437.2021.01350 https://doi.org/10.1023/b:visi.0000029664.99615.94

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2011–2023. https://doi.org/10.1109/tpami.2019.2913372

Hussein, F., Al-Ahmad, A., El-Salhi, S., Alshdaifat, E., & Al-Hami, M. (2022). Advances in Contextual Action Recognition: Automatic Cheating Detection Using Machine Learning Techniques. *Data*, *7*(9), 122. https://doi.org/10.3390/data7090122

Jackson, D. F. (1961). The Detection of Cheating in Multiple Choice Tests. *British Journal of Educational Psychology*, *31*, 54–64.

Jocher, G. A., & Qiu, J. (2023). *YOLO by Ultralytics*. https://github.com/ultralytics/ultralytics

Kaddoura, S., & Gumaei, A. (2022). Towards Effective and Efficient Online Exam Systems Using Deep Learning-Based Cheating Detection Approach. *Intelligent Systems with Applications*, *16*, 200153. https://doi.org/10.1016/j.iswa.2022.200153

Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). A Guide to Convolutional Neural Networks for Computer Vision. *Synthesis Lectures on Computer Vision*, *8*, 1–207.

Laurisa, S., M., F. O. K., & Tirtana, A. (2022). Cheating Detection During Exam with YOLO V4. *IC-ITECHS*, *3*(1), 231–235.

Li, Z., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., Xu, D., Ben, G., & Gao, Y. (2022). Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey. *Remote Sensing*, *14*(10), 2385. https://doi.org/10.3390/rs14102385

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327. https://doi.org/10.1109/tpami.2018.2858826

Lin, W., Y., Y. Z. X., Li, H. J., Dai, L., L, J. B. M., Zhao, H. W., & Chen, F. J. (2015). Cheating Behavior Detection in Examination Room Based on Background Subtraction. *Journal of Jilin University*, *29*, 406–409.

DOI: 10.3844/jcssp.2024.1668.1680

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Mittal, P., Singh, R., & Sharma, A. (2020). Deep Learning-Based Object Detection in Low-Altitude UAV Datasets: A Survey. *Image and Vision Computing*, *104*, 104046. https://doi.org/10.1016/j.imavis.2020.104046

Niu, Z., Zhong, G., & Yu, H. (2021). A Review on the Attention Mechanism of Deep Learning. *Neurocomputing*, *452*, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091

Pusara, M., & Pu, C. (2004). Dynamic Testing of Keystroke Dynamics for Identity Verification. *IFIP International Information Security Conference*, 390–401.

Qiao, S., S, T. H., Liu, G. H., & Wang, M. (2019). Object Detection Algorithm Based on Improved Feature Extraction Network. *Laser Optoelectron*, *56*(23), 134–139.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. https://doi.org/10.1109/cvpr.2016.91

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Resta, P., & Laferrière, T. (2008). *Issues and Challenges Related to Digital Equity* (J. Voogt & G. Knezek, Eds.; pp. 765–778). Springer US. https://doi.org/10.1007/978-0-387-73315-9_44

Wang, Q., Cheng, M., Huang, S., Cai, Z., Zhang, J., & Yuan, H. (2022). A Deep Learning Approach Incorporating YOLO v5 and Attention Mechanisms for Field Real-Time Detection of the Invasive Weed Solanum Rostratum Dunal Seedlings. *Computers and Electronics in Agriculture*, *199*, 107194. https://doi.org/10.1016/j.compag.2022.107194

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11531–11539. https://doi.org/10.1109/cvpr42600.2020.01155

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Vol. 11211, pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_1

Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent Advances in Deep Learning for Object Detection. *Neurocomputing*, *396*, 39–64. https://doi.org/10.1016/j.neucom.2020.01.085

Yu, L., Zhu, J., Zhao, Q., & Wang, Z. (2022). An Efficient YOLO Algorithm with an Attention Mechanism for Vision-Based Defect Inspection Deployed on FPGA. *Micromachines*, *13*(7), 1058. https://doi.org/10.3390/mi13071058

Yulita, I. N., Hariz, F. A., Suryana, I., & Prabuwono, A. S. (2023). Educational Innovation Faced with COVID-19: Deep Learning for Online Exam Cheating Detection. *Education Sciences*, *13*(2), 194. https://doi.org/10.3390/educsci13020194