Faculty of Computer Science and Information Technology

Improved Cheating Detection in Examinations using YOLOv8 with Attention Mechanism

Zuo Yan

Master of Science
2025

Improved Cheating Detection in Examinations using YOLOv8 with Attention Mechanism

Zuo Yan

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Computer Science)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2025

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

zuoyan

…………………………………

Signature

Name: Zuo Yan

Matric No.: 22020104

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date : 3th May, 2025

ACKNOWLEDGEMENT

I extend my deepest gratitude to my supervisor, Professor Soo See Chai, whose expertise, understanding, and patience, added considerably to my research experience. Your guidance and support have been invaluable throughout this journey. I am also immensely thankful to my committee members, for their insightful feedback and constructive critiques that significantly enhanced the quality of this work.

I am deeply grateful to my university and department for providing an enriching academic environment and the necessary resources. The financial support received from various grants and scholarships has been crucial in bringing this project to fruition.

I wish to acknowledge the contribution of my colleagues, whose perspectives and suggestions were enormously helpful. My heartfelt thanks go to the members, for their camaraderie, stimulating discussions, and all the fun we had during my time.

I must express my very profound gratitude to my parents, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

My friends deserve special mention for their unwavering support and constant encouragement throughout the duration of the project.

I would also like to acknowledge the professional support received from various individuals and organizations that contributed significantly to the quality of this work.

Finally, a sincere thank you to everyone who directly or indirectly contributed to the successful completion of this thesis.

ABSTRACT

Examinations are pivotal in educational and talent-selection processes globally. Ensuring their integrity is critical, but traditional invigilation methods, combining manual oversight with video monitoring, are resource-intensive and not fully effective in detecting cheating. This thesis presents an innovative approach to enhance examination fairness through automated cheating detection using an improved YOLOv8 algorithm integrated with an attention mechanism. The study focuses on developing and implementing a target- detection system capable of identifying student abnormal behaviors indicative of cheating in real-time. This system utilizes the advanced capabilities of the YOLOv8 model, optimized for speed and efficiency, making it suitable for deployment on regular- performance computers. The integration of an attention mechanism allows the system to focus on key visual cues that signify potential dishonesty, thereby improving detection accuracy. An original dataset, representing various cheating methods in paper-based exams, was manually compiled due to the absence of suitable open-source data. This dataset included a wide range of cheating behaviors, enabling comprehensive training and validation of the model. Results demonstrate that the improved YOLOv8 model achieves a detection accuracy of 82.71%, significantly reducing the need for manual video review and labor costs associated with traditional invigilation methods. This high accuracy rate meets the practical application requirements for real-time cheating detection in offline exam venues. This research contributes to the field of educational technology by offering a scalable, accurate, and efficient solution to a prevalent challenge in academic assessment. The successful implementation of this system can revolutionize examination monitoring, ensuring fairness and upholding academic integrity. Future work could extend this model's application to digital or oral examinations and explore its integration with other surveillance technologies. The proposed system's ability to function on standard computer hardware highlights its practical utility, while the incorporation of an attention mechanism ensures precise identification of cheating behavior amidst normal exam conditions. By addressing the limitations of existing invigilation practices, this study paves the way for more reliable and equitable examination environments. Ultimately, the adoption of such advanced technologies could establish new standards in academic integrity, benefiting educational institutions and students alike.

Keywords: Examinations, student abnormal behavior, detection, improved YOLOv8, attention mechanism, cheating detection

Peningkatan Pengesanan Meniru dalam Peperiksaan menggunakan YOLOv8 dengan Mekanisme Perhatian

ABSTRAK

Peperiksaan adalah penting dalam proses pendidikan dan pemilihan bakat di seluruh dunia. Memastikan integriti mereka adalah kritikal, tetapi kaedah pengawasan tradisional, menggabungkan pengawasan manual dengan pemantauan video, adalah intensif sumber dan tidak berkesan sepenuhnya dalam mengesan penipuan. Tesis ini membentangkan pendekatan inovatif untuk meningkatkan keadilan peperiksaan melalui pengesanan penipuan automatik menggunakan algoritma YOLOv8 yang dipertingkatkan yang disepadukan dengan mekanisme perhatian. Kajian ini memberi tumpuan kepada pembangunan dan pelaksanaan sistem pengesanan sasaran yang mampu mengenal pasti tingkah laku abnormal pelajar yang menunjukkan penipuan dalam masa nyata. Sistem ini menggunakan keupayaan canggih model YOLOv8, dioptimumkan untuk kelajuan dan kecekapan, menjadikannya sesuai untuk penggunaan pada komputer berprestasi biasa. Penyepaduan mekanisme perhatian membolehkan sistem memfokus pada isyarat visual utama yang menandakan potensi ketidakjujuran. Dengan itu, ketepatan pengesanan dapat ditingkatkan. Set data asal, yang merangkumi pelbagai kaedah penipuan dalam peperiksaan berasaskan kertas, telah disusun secara manual disebabkan ketiadaan data sumber terbuka yang sesuai. Set data ini termasuk pelbagai tingkah laku menipu, membolehkan latihan komprehensif dan pengesahan model Keputusan menunjukkan bahawa model YOLOv8 yang dipertingkatkan mencapai ketepatan pengesanan sebanyak 82.71%, dengan ketara mengurangkan keperluan untuk semakan video manual dan kos buruh yang dikaitkan dengan kaedah pengawasan tradisional. Kadar ketepatan yang tinggi ini memenuhi keperluan aplikasi praktikal untuk pengesanan penipuan masa nyata di tempat *peperiksaan luar talian. Penyelidikan ini menyumbang kepada bidang teknologi pendidikan dengan menawarkan penyelesaian berskala, tepat dan cekap kepada cabaran lazim dalam penilaian akademik. Kejayaan pelaksanaan sistem ini boleh merevolusikan pemantauan peperiksaan, memastikan keadilan dan menegakkan integriti akademik. Kerja masa hadapan boleh memanjangkan aplikasi model ini kepada peperiksaan digital atau lisan dan meneroka integrasinya dengan teknologi pengawasan lain. Keupayaan sistem yang dicadangkan untuk berfungsi pada perkakasan komputer standard menyerlahkan utiliti praktikalnya, manakala penggabungan mekanisme perhatian memastikan pengecaman tepat tingkah laku menipu dalam keadaan peperiksaan biasa. Dengan menangani batasan amalan pengawas sedia ada, kajian ini membuka jalan kepada persekitaran peperiksaan yang lebih dipercayai dan saksama. Akhirnya, penggunaan teknologi canggih tersebut boleh mewujudkan piawaian baharu dalam integriti akademik, memberi manfaat kepada institusi pendidikan dan juga pelajar.*

Kata kunci: *Peperiksaan, tingkah laku abnormal pelajar, pengesanan, peningkatan YOLOv8, mekanisme perhatian, pengesanan penipuan*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

YOLO	You Only Look Once
CNNs	convolutional neural networks
SVM	support vector machines
R-CNN	region-based convolutional neural network
SPPNet	spatial pyramid pooling network
SSD	single-shot multibox detector
FC	fully connected
AM	attention mechanisms
SE	squeeze-and-excitation
CBAM	convolutional block attention module
ECA	efficient channel attention
CA	coordinate attention
SOCA	second-order attention network

CHAPTER 1

INTRODUCTION

1.1 Research Background

The fairness and integrity of examinations are critical to maintaining trust in educational assessments. However, the increasing prevalence of cheating in both traditional and digital examination environments has raised significant concerns for educators and institutions worldwide. The transition to online and hybrid examination formats, driven by the advancement of educational technology, has further complicated the issue. The lack of physical invigilation in such formats has made cheating easier and more sophisticated, highlighting the urgent need for robust and effective cheating detection systems (Altbach et al., 2019). Studies indicate that despite debates around their relevance, the global demand for examinations has remained steady, particularly in regions where standardized testing is tied to academic progression and career opportunities (Morris et al., 2021). Academic institutions are now prioritizing the development and adoption of advanced technologies to detect cheating during examinations. Universities and schools are actively integrating electronic proctoring tools and AI-based surveillance systems to ensure secure and fair examination environments (Sindre et al., 2015). These systems are designed to identify various forms of cheating, such as unauthorized material usage, suspicious movements, and inappropriate communication, thereby upholding the credibility of assessment processes.

Dishonesty during examinations undermines not only individual academic performance but also the overall credibility of education systems. From universities to secondary and even primary schools, the rise in cheating practices has necessitated the implementation of innovative solutions. While some institutions had planned to adopt such technologies before the pandemic (Seo, 2013), the urgency of addressing these challenges has intensified in recent years (Krishnamurthy, 2020). Advanced AI-powered frameworks, such as improved object detection models, are being explored to monitor examination environments in real time and detect subtle behaviors indicative of cheating. These technologies play a vital role in ensuring the integrity of examinations, safeguarding the trust of educators, students, and stakeholders in the fairness of assessment systems.

The detection of cheating in conventional examination settings was typically accomplished through human proctors, or more recently, through technology-assisted solutions. A common approach involves the use of video surveillance systems (Hayes, 2007; Pandita et al., 2023), in which proctors monitor examinees through live streams or recorded video footage. However, these systems still rely heavily on human vigilance, which can be flawed by fatigue, inattentiveness, or bias (Jackson, 1961; Huang et al., 2023). Some systems have integrated computer-based technologies, such as keystroke dynamics analysis (Pusara et al., 2004; Perera et al., 2023), in which unique typing patterns are used to verify the identity of the test-taker. Nonetheless, this approach is restricted to computer-based examinations and does not directly address acts of cheating during an examination. Biometric systems using facial or iris recognition have also been utilized (Schockaert et al., 2010; Lee et al., 2023); however, these methods may encounter technical glitches, such as poor lighting conditions (Ma et al., 2023), or may raise privacy concerns (Michael et al., 2014; Wenhua et al., 2023).

The use of automatic processing software for cheating detection has grown in prominence, particularly with the increase in examinations. Such software uses machine-learning techniques to analyze student behaviors during the examination of potential cheating indicators (Kumar et al., 2017). Nevertheless, these programs often generate false positives and negatives because of their inability to accurately interpret all human behaviors and the context for such behaviors (Resta et al., 2018). To address these limitations, several studies have proposed the use of machine-learning algorithms, such as support vector machines (SVM) and neural networks, to detect cheating (Chen et al., 2019; Zhou et al., 2023). Although these techniques can process large amounts of data and learn complex patterns, they may lack interpretability and require substantial labelled data for effective functioning (Ribeiro et al., 2017; Taye, 2023). Recognizing these shortcomings, researchers have explored more specialized techniques such as object detection, which can offer greater precision in identifying specific cheating-related objects or actions within the examination environment. The use of object-detection algorithms, such as You Only Look Once (YOLO), presents a promising avenue for cheating detection. YOLO, in particular, is a state-of-the-art real-time object-detection system that identifies objects in a given image or video sequence (Redmon et al., 2016). However, earlier versions of YOLO could not focus on small or ambiguous objects, the detection of which could be crucial for detecting subtle cheating behaviors (Khan et al., 2018). Therefore, this study proposes a new YOLOv8 algorithm comprising an attention mechanism with the potential to address these limitations. The attention mechanism allows the model to focus on important areas in an image or video, potentially enhancing its ability to detect minor cheating behaviors.

This study focuses on identifying the current action labels based on video sequences to detect cheating during examinations. In addition, a comprehensive methodology that can detect and classify anomalous behaviors and actions occurring within a test room is presented, potentially indicating instances of cheating. This methodology was developed based on YOLOv8 and the attention mechanism. Due to the lack of an open-source dataset specifically created for detecting cheating in paper-based exams, a dataset was manually constructed and compiled. This dataset was developed to illustrate the various methods students might use to deceive a proctor during a paper-based examination. The dataset comprises the most frequent cheating techniques, which are logically categorized into seven types: normal, passing items (e.g., notes, rulers), whispering, putting hands under the table, taking out mobile phones, walking around, and looking around. These seven categories were selected based on their practicality and relevance in traditional paper-based examinations. They represent a comprehensive yet manageable framework of common cheating behaviors observed in exam settings. The classification ensures coverage of both direct cheating methods (e.g., passing items or using mobile phones) and indirect methods (e.g., whispering or looking around). Such classifications align with research on academic dishonesty, which identifies diverse methods used by students, including exchanging materials, using technology, non-verbal signals, and movement during exams (Faucher & Caves, 2009). Whispering and subtle body gestures are validated as common tactics in exams (Asadullah & Nisar, 2016). The use of mobile phones and technology for cheating reflects the increasing sophistication of academic deceit (Lipson & Karthikeyan, 2016). Observational behaviors such as looking around and walking are noted as traditional yet persistent methods of dishonesty (Odongo et al., 2021).

1.2 Problem Statement

Examinations serve as a cornerstone of academic and professional assessment frameworks globally. The integrity and fairness of examinations are pivotal to ensuring the validity of assessment outcomes, which in turn influence academic credentials and career trajectories (Zheng et al., 2021). Despite the critical role of examinations, the traditional approach to invigilation has been predominantly manual, relying on the presence of invigilators and the use of surveillance technologies such as closed-circuit television

(CCTV). This method, while widely practised, poses several challenges (Redmon et al., 2018). Firstly, manual invigilation is resource-intensive, demanding significant human involvement, which may lead to subjective judgments and oversight due to human error or fatigue. Additionally, the sheer volume of video data generated during examinations necessitates a substantial investment in time and labor to review, often resulting in delayed or ineffective detection of malpractices. Moreover, manual reviews do not guarantee comprehensive coverage or real-time detection of cheating, thereby compromising the fairness and credibility of the examination process (Davis, 2019).

There is an urgent need to enhance the efficiency and efficacy of cheating detection during exams. The use of automated systems for monitoring and detection purposes has been explored to address these challenges. However, existing automated surveillance systems have limitations in terms of accuracy, speed, and computational demands, which impede their practicality for real-time application in diverse examination settings (Bochkovskiy et al., 2020). Another significant challenge lies in the technical computational limitations of existing automated surveillance systems. While automated systems offer potential advantages in terms of efficiency and scalability, they often struggle with issues such as low detection accuracy, high computational complexity, and latency in real-time scenarios. Many current models rely on extensive computational resources, making them impractical for deployment in resource-constrained environments such as educational institutions with limited budgets (Kornaros, 2022). Furthermore, these systems may fail to adapt effectively to the diversity of cheating behaviors due to insufficient generalization capabilities, particularly when dealing with novel or less predictable patterns. For instance, some existing models suffer from high false positive and false negative rates, which undermine their reliability and utility. Additionally, the processing of large-scale video data in real-time requires both speed and precision, yet many existing frameworks exhibit bottlenecks in terms of computational efficiency, leading to delayed or incomplete detection (Li et al., 2020). These limitations underscore the necessity for developing a solution that not only achieves high accuracy but also operates efficiently within the constraints of real-time application and diverse examination settings.

The study at hand seeks to address these challenges by proposing an improved YOLO model, specifically YOLOv8, integrated with an attention mechanism designed to detect cheating behaviors accurately and in real time. The introduction of the attention mechanism aims to refine the detection process by focusing on relevant features and patterns indicative of cheating, thereby enhancing detection accuracy (Lin et al., 2017). The YOLOv8 model's improved speed and computational efficiency make it feasible for deployment in standard examination settings without the need for high-end computing resources. The escalating prevalence of dishonest behaviors in various settings necessitates the development of more sophisticated detection methods. Traditional surveillance and monitoring techniques often fall short of accurately identifying subtle cues and patterns that are indicative of cheating. This gap highlights the urgent need for advanced technological solutions capable of adapting to the complexity of human behavior. The incorporation of attention mechanisms into the YOLOv8 framework represents a significant step forward in this direction. Incorporating attention mechanisms into the YOLOv8 framework is beneficial because these mechanisms enable models to focus on relevant features within vast datasets. By selectively concentrating on the most pertinent parts of the input, attention mechanisms enhance the model's ability to discern intricate patterns and nuances that might otherwise be overlooked. This improved focus leads to more accurate identification of subtle indicators of cheating, thereby making the surveillance system more effective and reliable. This research objective stems from the recognition that existing methods lack the precision and adaptiveness required to effectively counteract cheating. By augmenting the YOLOv8 model with attention mechanisms, this study aims to bridge this gap, offering a more accurate, efficient, and reliable means of detecting dishonest behaviors.

The research questions guiding this study are:

(a) How can the YOLOv8 model be optimized to detect specific cheating behaviors in an examination setting?

(b) What is the effectiveness of incorporating an attention mechanism into the YOLOv8 framework in improving the model's ability to detect subtle cues and patterns indicative of cheating?

(c) Can the proposed system operate effectively in real-time on standard-performance computers?

The scope of this research encompasses various forms of cheating behaviors that can be captured on video within examination rooms, including but not limited to unauthorized material usage, irregular movement, and communication with peers (Vaswani et al., 2017). These behaviors align with the dataset described in Section 1.1, which classifies the most frequent cheating techniques into seven distinct types: normal, passing items (e.g., notes, rulers), whispering, placing hands under the table, taking out mobile phones, walking around, and looking around. The behaviors analyzed in this study are drawn directly from these classifications, ensuring that the scope is consistent with the dataset's defined categories. By focusing on these specific types, the research aims to develop a model capable of detecting subtle and overt cheating behaviors effectively in real-time. The study's parameters include the evaluation of the model's detection accuracy, speed, and adaptability to participant behaviors. These parameters are designed to comprehensively assess the model's effectiveness and reliability in detecting cheating behaviors under realistic and varied conditions.

The significance of this research lies in its potential to revolutionize the field of examination invigilation by introducing a scalable, accurate, and efficient automated system for cheating detection. Such a system could potentially save educational institutions significant amounts of time and resources while upholding the integrity of examinations. It would provide a more equitable assessment atmosphere where all students are subject to the same rigorous monitoring standards.

By improving the standard of monitoring and detection, the proposed system also aims to act as a deterrent to potential malpractices, thereby fostering a culture of honesty and integrity within academic environments. Furthermore, the findings from this study could be applicable to other surveillance and monitoring domains where real-time detection of specific behaviors is critical.

In conclusion, the proposed study addresses a gap in the current examination invigilation methodology by offering an automated, intelligent system capable of enhancing the detection of cheating behaviors. This system not only promises to improve the fairness and integrity of examinations but also serves as a model for integrating advanced machine learning techniques into practical applications.

1.3 Research Objectives

To address the research questions highlighted in previous section, the following research objectives are outlined:

Adapt and customize YOLOv8 for cheating detection: This objective is to adapt and customize the YOLOv8 algorithm to specifically identify various cheating behaviors in examination settings. This includes enhancing the model's ability to accurately detect actions such as the use of unauthorized materials, unusual movements, and covert communication among students. The customization process will focus on adapting the algorithm to be moreresponsive and effective in recognizing these specific types of misconduct during exams.

To incorporate an attention mechanism into the YOLOv8 framework with the objective of enhancing the model's ability to detect subtle cues and patterns indicative of cheating. This integration aims to test and evaluate how attention mechanisms can improve the accuracy and efficiency of the model in identifying dishonest behavior.

Optimization for Real-Time Application: The objective here is to ensure the modified YOLOv8 model operates efficiently in real-time on standard- performance computers. This involves optimizing the algorithm to process and analyze video feeds swiftly, enabling immediate detection and response to cheating incidents in various exam environments.

1.4 Research Scope

The focus of this research is specifically on the detection and classification of cheating behaviors in paper-based examination settings through the analysis of video surveillance data. This scope is deliberately narrow to ensure a detailed and targeted investigation into how video analytics can be effectively utilized to identify instances of academic dishonesty during exams. To establish a relevant action label system, this study categorizes cheating behaviors into seven distinct types, as follows:

1. Normal Behavior: Actions that represent non-cheating-related activities, such as writing, reading, or sitting still. These serve as a baseline to improve detection accuracy by distinguishing normal behaviors from anomalous ones.

2. Passing Items: The transfer of unauthorized materials, such as notes, rulers, or other objects, between students. This behavior is explicitly prohibited in examination settings.

3. Whispering: Any oral communication or attempted communication between students, which may involve sharing answers or other restricted information.

4. Placing Hands Under the Table: Hidden hand movements, such as placing hands beneath the table, which may indicate the student is manipulating unauthorized materials out of sight.

5. Taking Out Mobile Phones: The act of retrieving or using a mobile phone during an examination. This behavior is a clear violation of exam protocols in paper-based examination environments.

6. Walking Around: Unpermitted movement within the examination room, such as walking without valid reasons, which could suggest attempts to gain unauthorized access to information or collaborate with others.

7. Looking Around: Wandering gaze or excessive head movements indicating that the student may be attempting to observe other students' answers or find external cues.

To facilitate this study, a specialized dataset is manually curated, reflecting the nuanced spectrum of cheating methods in examination settings. This dataset is a cornerstone of the research, given the lack of existing open-source datasets tailored to the context of cheating detection in paper-based exams. It encompasses a variety of scenarios that illustrate both overt and subtle forms of cheating, as well as baseline 'normal' behavior to provide contrast for the machine learning model.

The framework developed employs the YOLOv8 algorithm, known for its proficiency in real-time object detection, which has been enhanced with an attention mechanism to pinpoint features within the video that are characteristic of cheating. The attention mechanism serves to focus the model on relevant spatial and temporal features, thus refining the detection process.

The research's applicability is intended for real-world examination settings, proposing a system that not only automates the detection process but also introduces a level of precision and reliability not currently present in manual invigilation methods. Moreover, the development and validation of this framework involve rigorous testing under controlled conditions to simulate a variety of examination environments.

Although the research is comprehensive, it acknowledges certain limitations. For instance, it focuses exclusively on paper-based examination scenarios and does not cover digital or oral examinations. Furthermore, the current model is designed to operate in controlled environments with predetermined camera positions and clear visibility of the test-takers. The methodology's efficacy in different lighting conditions, camera qualities, and angles remains to be thoroughly evaluated.

In addition to the technical aspects, the research will address ethical considerations regarding the use of surveillance technology in educational settings, including privacy, data security, and the implications of false positives in cheating detection.

By delineating these boundaries, the research aims to establish a focused and effective approach to detecting cheating, providing educational institutions with a valuable tool to uphold academic integrity and ensure fair assessment of student performance. The creation of a proprietary dataset also opens avenues for future research, where subsequent studies might expand upon the types and conditions of cheating behaviors analyzed or adapt the framework for use in alternative testing formats and environments.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The proliferation of technology in educational settings has brought to light the challenges and opportunities within the assessment ecosystem. This literature review explores the body of knowledge surrounding the technologies and methodologies pertinent to this cause. Beginning with an overview of computer vision, this section will delve into the intricacies of CNNs and their relevance in object detection tasks. This paper further dissect the evolution and application of object detection algorithms, culminating in their application in the niche but crucial area of cheating detection during examinations. Lastly, the role of comprehensive datasets in training and validating these detection systems is discussed, highlighting the importance of targeted and well-annotated data in the development of robust models. Through this review, this paper aim to provide a thorough understanding of the current state of technology and identify gaps that our research addresses, setting the stage for the ensuing chapters that detail our methodological advancements and empirical contributions.

2.2 Computer Vision in Examination Settings

Computer vision has evolved as an indispensable aspect of AI, offering the capability to endow machines with the power to interpret and understand the visual world. Drawing parallels to the human visual system, computer vision algorithms enable computers to identify and process objects in images and videos in a way that mimics human recognition (Mitra et al., 2022). The relevance of computer vision in the domain of cheating detection during examinations is profound, as it integrates the sophistication of AI in the monitoring process to ensure the integrity and fairness of assessments (Kontogianni et al., 2022).

The proliferation of computer vision in examination settings is propelled by the need to reduce the reliance on human invigilation, which is labor-intensive and susceptible to oversight. Given the stringent requirements for maintaining exam decorum, computer vision systems can assist invigilators by continuously analyzing video feeds to detect possible instances of misconduct (Jiang et al., 2022). This technology integration leverages the developments in various AI disciplines, including pattern recognition, machine learning, and image processing, to create a robust surveillance mechanism that can operate with minimal human intervention.

In the context of cheating detection, the role of computer vision is to provide a relentless and unblinking eye that can oversee multiple candidates simultaneously, with a level of diligence far beyond human capabilities (Lamping et al., 2022). For instance, the use of computer vision allows for the detection of subtle movements or gestures that may indicate cheating, such as glancing at a neighbor's paper, unauthorized use of electronic devices, or even signaling to a peer. By processing the visual data, the system can flag these actions in real-time, allowing for immediate intervention by exam proctors.

To effectively integrate computer vision into exam supervision, it is essential to understand the underlying technologies that enable this application. Among these, Convolutional Neural Networks (CNNs) are particularly noteworthy. As a type of deep learning algorithm optimized for processing visual data, CNNs are well-suited for tasks such as image recognition and classification, making them ideal for monitoring exam environments. These networks analyze visual inputs through multiple layers that detect an interpret various patterns, shapes, and textures. This capability is crucial for identifying potential cheating behaviors or unauthorized materials during exams. A practical application of this technology involves using CNNs to analyze video sequences that depict various cheating behaviors, which are often simulated by actors for training purposes. This approach allows the system to learn and classify different cheating actions. For instance, Hussein et al. (2022) demonstrated a system that performs action recognition at the frame level, using well-known features to achieve an accuracy of 82% in detecting cheating behaviors. The integration of CNNs in exam supervision involves complex challenges, as highlighted by Wise et al. (2022). The primary issue identified in their research is the algorithm's capacity to accurately differentiate between legitimate exam-taking behaviors and fraudulent activities. This distinction is critical, as CNNs must be trained to discern subtle nuances between normal and cheating behaviors within the confines of an exam room. Developing such sophisticated models requires an in-depth understanding of typical behaviors during examinations, as well as a comprehensive catalog of potential cheating tactics.

According to Wise et al. (2022), the design of these models begins with defining what constitutes normal behavior in an exam setting. Normal behaviors might include students looking down at their own papers, quietly thinking, or occasionally stretching, which should not trigger any alarms within the supervisory system. In contrast, examples of potential cheating behaviors identified in their research include frequent glances at other students' papers, unauthorized use of electronic devices, or suspicious body language that suggests exchanging information. The challenge lies in programming CNNs to recognize and interpret these behaviors accurately, thus ensuring fairness and integrity in the exam supervision process. The application of CNNs in this scenario heavily relies on their detection and classification abilities, which are adept at identifying patterns within visual data. As noted by Gudmundsson et al. (2022), CNNs are particularly effective for image recognition tasks due to their hierarchical architecture. This structure enables them to learn complex features from raw pixel data, progressively recognizing various objects' shapes, sizes, and textures, which is critical for identifying materials or actions associated with cheating.

However, in the context of monitoring the dynamic and unpredictable environment of an examination hall, traditional CNNs may encounter significant challenges. Traditional CNNs, primarily designed for static image analysis, struggle with real-time processing and the dynamic nature of video streams, where exam conditions and student behaviors are continuously changing (Sharma et al., 2021). They also often require large amounts of data for training to recognize complex behaviors accurately, which can be difficult to gather in the nuanced and varied context of an exam room. Additionally, the fixed architecture of traditional CNNs may not be flexible enough to adapt to the diverse array of possible cheating methods and innocent yet suspicious-looking behaviors.

YOLO, or "You Only Look Once," (Jocher et al., 2023) represents a significant evolution from conventional CNNs by integrating the detection process into a single evaluation step. This innovative approach enables YOLOv8 to process images at unparalleled speeds, making it an ideal solution for the real-time detection of cheating during examinations. Its ability to analyze video feeds with minimal delay is critical, as it ensures that any act of cheating can be immediately identified and addressed, maintaining the integrity of the examination process. YOLOv8 is the latest iteration in the YOLO series, which began with the original YOLO algorithm and has seen several improvements over the years, including YOLOv2 (YOLO9000), YOLOv3, and YOLOv4. Each version has introduced enhancements in speed, accuracy, and the ability to detect smaller objects, culminating in the highly efficient YOLOv8 used in this study.

The detection accuracy of YOLOv8 was improved by coupling attention mechanisms, including wrist fracture detection, elevator passenger detection, distracted driving behavior and driver's emotion detection, fruit freshness detection, road defect. (Chien et al., 2024; Wang et al., 2024 ; Ma et al., 2024; Wei et al., 2023; Wang et al. 2023). The attention mechanism is an AI technique that mimics cognitive attention, enabling the model to focus on specific parts of the visual data that are more relevant for the task at hand. In the realm of cheating detection, this means prioritizing regions in the exam room where cheating is more likely to occur, such as students' desks and their immediate vicinity. By focusing on these areas, the improved YOLOv8 will yield higher accuracy in detecting irregular activities indicative of cheating.

This technology promises a new era of efficiency and effectiveness in maintaining the sanctity of examinations. By reducing the manual labor associated with invigilation and increasing the detection of fraudulent activities, such systems help uphold fairness in academic evaluations. As computer vision technology continues to advance, its potential to transform the landscape of examination invigilation grows, signaling a move towards more secure and equitable educational practices.

2.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) represent a paradigm shift in visual recognition that has revolutionized the field of computer vision, especially in tasks requiring the categorization and detection of objects within images (Kim et al., 2014,Howard et al., 2017, Tara et al., 2015). Their architecture is inspired by the biological processes of the human brain, particularly the organization of the animal visual cortex. These networks are highly effective in extracting hierarchical features from visual data, which is essential for a system designed to detect cheating in examinations using video data.

A CNN typically consists of a series of layers that transform the input image into a form that can be processed to make a prediction or classify an image (Jaderberg et al., 2014). The layers are composed of a series of convolutional layers, pooling layers, and fully connected layers at the end. The convolutional layer is the core building block of a CNN and is responsible for the detection of various features in the image. It uses a set of learnable filters that apply convolution operations to the input. Each filter in a convolutional layer is small spatially but extends through the full depth of the input volume. For a given filter W of size k×k×d applied to an input volume X with stride S, the convolution operation is defined as:

$$Y_c(x,y) = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{l=1}^{d} W(i,j,l) \cdot X(x+i-1, y+j-1, l) \qquad \text{Equation 2.1}$$

where $Y_c(x,y)$ is the output of the convolution operation at position (x,y) on the feature map corresponding to the c-th filter.

Pooling layers follow the convolutional layers and serve to reduce the spatial dimensions of the representation, which decreases the number of parameters and computation in the network, thereby controlling overfitting. The most common form of pooling is max pooling, which takes the maximum value over a set of spatial locations. If we define P as the pooling size and Z as the input to the pooling layer, the max pooling operation over a window P at location (x,y) can be expressed as:

$$M(x, y) = \max_{a=0}^{P-1} \max_{b=0}^{P-1} Z(x \cdot S + a, y \cdot S + b) \qquad \text{Equation 2.2}$$

where S is the stride of the pooling operation, and $M(x, y)$ is the output after the pooling operation.

Fully connected layers are the last stage of a CNN. Each neuron in a fully connected layer has full connections to all activations in the previous layer, as seen in regular Neural Networks. Their role is to take the high-level features extracted by the convolutional and pooling layers to make a final prediction. For a fully connected layer with input vector a, weights W, and bias b, the output o is computed as:

$$\mathrm{o} = \mathrm{W} \cdot \mathrm{a} + \mathrm{b} \qquad \text{Equation 2.3}$$

The non-linear activation function, often a Rectified Linear Unit (ReLU), introduces non-linear properties to the system, allowing the network to learn complex mappings from data. Without such non-linearity, the CNN would be unable to learn and model more complex forms of data, such as the nuances of cheating behavior in examination settings.

Training a CNN involves the adjustment of weights in all filters across the network's layers to minimize the error in prediction. The backpropagation algorithm is used in conjunction with an optimization algorithm such as stochastic gradient descent (SGD) to facilitate this learning process. SGD updates the network's weights iteratively based on the gradient of the loss function L with respect to the weights W, defined by the equation:

$$W_{\text{new}} = W_{\text{old}} - \eta \nabla_W L(W_{\text{old}})$$

where η is the learning rate, and $\nabla_W L(W_{\text{old}})$ is the gradient of the loss function concerning the weights.

Figure 2.1：shows an example of an aircraft structural health was detected via a CNN.The process begins with an aircraft sensing input, which is an image or a set of sensor data representing the aircraft's condition. The first stage of the CNN consists of convolution layers paired with ReLU (Rectified Linear Unit) activation functions. These layers are responsible for feature extraction by applying various filters that detect edges, textures, and other relevant features within the input data.

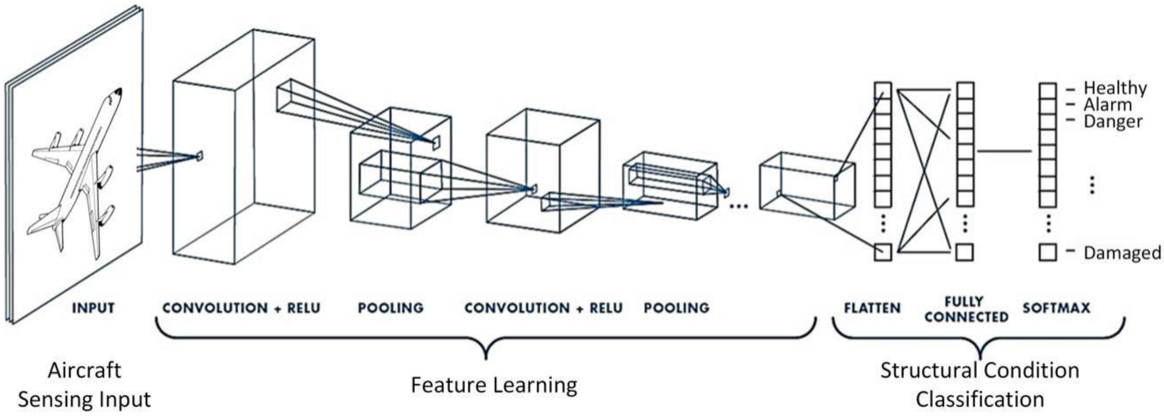

Figure 2.1: Schematic diagram of how the convolutional layer works to detect (Tabian et al., 2019)

Subsequent pooling layers downsample the feature maps to reduce dimensionality, which aids in making the network less sensitive to the exact location of features in the input space and decreases the computational load. The combination of convolution and pooling layers effectively captures the hierarchy of features - from simple to complex - essential for discerning the aircraft's structural integrity.

After several convolution and pooling layers, the feature maps are flattened into a single vector, which is then fed into a fully connected (FC) layer. This dense layer integrates the learned high-level features to make decisions about the data. The final layer is a softmax layer, which outputs the probabilities of different classes. In this context, the classes are likely conditions of the aircraft's structure, such as 'Healthy', 'Danger', or 'Damaged'. The softmax layer ensures that the output probabilities sum to one, allowing for a clear and interpretable classification.

In the specific context of cheating detection in examinations, the design of the CNN must consider the variability and complexity of the examination environment. The network must be capable of capturing a wide range of features associated with cheating, such as the orientation of students' heads, eye gaze, hand movements, and the presence of unauthorized materials. This requirement necessitates the network to have a deep architecture with multiple layers to extract low, mid, and high-level features that comprehensively represent the examination setting (Su et al., 2015, Dong, et al., 2015).

Kohli et al. (2022) discussed the use of deep learning, particularly CNNs, in recognizing gestures indicative of cheating behaviors during exams. The study highlights the automatic extraction of spatiotemporal features from raw video input, which is a significant advancement over traditional methods. The CNN model, initially used for static image processing, has been adapted to handle sequential video frames as multi-channel inputs, allowing for effective video classification. Sharma et al. (2021) highlighted the broader context of exam cheating and the challenges posed by learning environments, especially during the COVID-19 pandemic. It underscores the need for effective technological solutions, such as CNNs, to address these challenges. As can be seen, CNN's impact on cheating detection systems is significant, providing a technical foundation for creating a more secure and fair examination environment. The deployment of such systems could potentially transform the landscape of educational assessments, where the demand for reliable, automated supervision is growing (Zhou et al., 2015). As the technology matures, it is poised to become an invaluable tool in preserving academic integrity and reinforcing the credibility of educational credentials.

2.4 Object Detection Algorithms

Building upon the advancements in CNNs discussed in Section 2.2, Object Detection Algorithms emerge as a critical component in the realm of computer vision, particularly in detecting and localizing objects within images or video sequences (Padilla et al., 2020). These algorithms are essential in identifying cheating behaviors in exams, serving as a practical application of the CNN advancements. Object detection algorithms can be categorized into two types: one-stage detectors, which combine classification and localization in a single process (Mariano et al., 2002, Zhao et al., 2017), and two-stage detectors, which initially generate regions of interest and subsequently classify these regions (Hu et al., 2017; Carranza, et al., 2020). The evolution and efficacy of these algorithms are inextricably linked to the progress in CNNs. Their role in automated examination invigilation is pivotal, as they provide the necessary tools for effective cheating behavior detection. The convergence of CNNs and object detection algorithms exemplifies how advancements in one area of computer vision can significantly enhance the capabilities in another, particularly in the context of maintaining integrity in educational assessments.

One-stage detectors such as YOLO and Single Shot MultiBox Detector (SSD) (Liu et al., 2016) prioritize inference speed, which is a crucial factor for real-time applications such as cheating detection during live exams. YOLO divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell. It treats the detection as a regression problem, directly predicting the bounding box coordinates and class probabilities with a single forward pass of the network. This method drastically reduces the processing time, making it suitable for real-time detection tasks. SSD extends this concept by using multiple feature maps at different scales to perform detections, allowing for more accurate detection of objects of various sizes (Qu et al., 2020).

Two-stage detectors, such as the R-CNN (Region-based Convolutional Neural Networks) family (He et al., 2017), including Fast R-CNN, Faster R-CNN, and Mask R-CNN, are known for their high accuracy. These models work by first proposing candidate object bounding boxes and then classifying each box using a convolutional neural network. For instance, the R-CNN algorithm applies a selective search to generate potential bounding boxes in the image and then uses a CNN to classify the objects within each box. Fast R-CNN improved on this by sharing the computation of the CNN across the different region proposals (Yadav et al., 2017). Faster R-CNN further enhances the process by introducing a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus speeding up the process and improving accuracy. Mask R-CNN extends Faster R-CNN by adding a branch for predicting segmentation masks on each Region of Interest (RoI), which is useful for differentiating objects at the pixel level, an essential feature for detailed analysis required in monitoring examinations (Du et al., 2020, Cuevas et al., 2016).

The application of the aforementioned object detection algorithms, including both one-stage and two-stage approaches, in the field of examination invigilation offers numerous advantages (Nascimento et al., 2006). These algorithms can process vast amounts of video data in real- time, providing immediate alerts to suspicious activities. For example, by training a model on a dataset of cheating behaviors, an object detection system can learn to recognize the presence of unauthorized materials, such as notes or mobile phones, and subtle gestures, such as an unusual head movement or eye gaze, which could suggest that a candidate is looking at another's work or engaging in illicit communication.

However, the challenge in applying object detection to cheating detection lies in the diversity and subtlety of cheating behaviors, which are not always overt or easily distinguishable from permissible behavior (Arandjelović et al., 2019). This has necessitated the development of sophisticated objective detection models that not only recognize a wide range of objects but also contextualize their usage within the exam environment.

For object detection algorithms, the detection process can be conceptualized as follows: an image from a video feed is input into a CNN, which extracts features at various scales. The network applies an attention mechanism to prioritize certain features over others, enhancing the detection of specific behaviors (Lu et al., 2019). It then predicts bounding boxes and assigns confidence scores that indicate the likelihood of a bounding box containing a particular type of cheating behavior. These predictions are subsequently filtered through a thresholding step to determine the final detections.

An essential part of this process is non-maximum suppression (NMS), which resolves the issue of multiple bounding boxes being predicted for the same object. NMS retains only the box with the highest confidence score, thus reducing redundancy and improving the accuracy of detection (Jiang et al., 2018). In the context of examination cheating detection, NMS is pivotal for isolating distinct cheating behaviors. For example, the algorithm has been applied to detect multiple types of cheating behaviors such as looking at hidden notes, signaling to peers, and unauthorized use of electronic devices.

The mathematical formulation for NMS is critical to understand its function in the detection process given a set of predicted bounding boxes B and corresponding confidence scores S, for each box B_i, the algorithm first selects the box B_j with the highest score S_j. It then compares B_j with all other boxes Bk and calculates the Intersection over Union (IoU).

If IoU between B_j and Bk exceeds a predefined threshold, Bk is suppressed. The output of the NMS process is a set of bounding boxes that have been pruned to include only the highest scoring boxes without significant overlap (Sun et al., 2021). The Figure 2.2 shows an example of NMS.

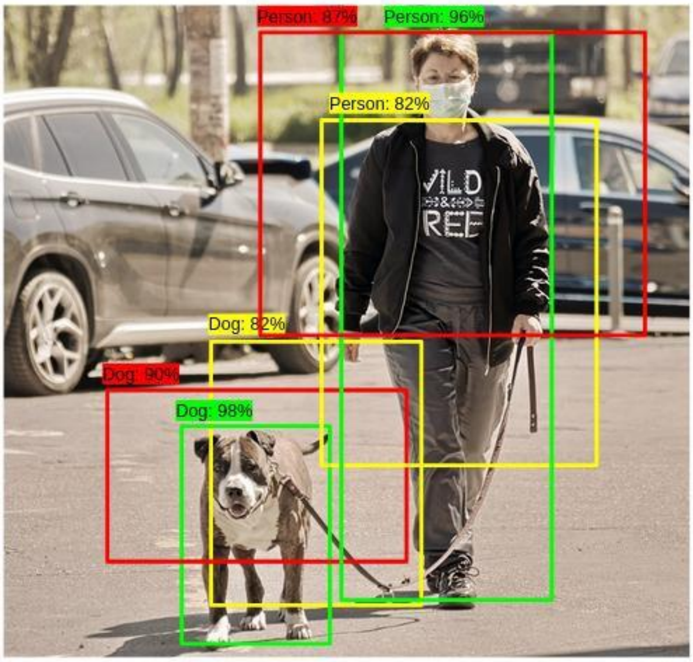

Figure 2.2: Right Bounding Box Using Non-Max Suppression

In conclusion, object detection algorithms form an integral part of the technology needed to monitor and maintain the integrity of examination processes. With the ability to analyze video feeds and detect cheating in real-time, these algorithms are pivotal in automating the invigilation process, thereby saving costs, resources, and ensuring the fairness and credibility of the examination system. As these algorithms continue to evolve, they hold the promise of creating an increasingly secure and unbiased assessment environment.

2.5 Cheating Detection

Cheating detection in examination settings is a critical issue that educational institutions must address to ensure the integrity and fairness of the assessment process. With the advent of sophisticated surveillance technologies, the potential for employing automated systems to detect cheating has become a focal point for research and development. Among various object detection methods, YOLO has emerged as a highly effective framework due to its speed and accuracy, which are vital in monitoring and detecting cheating during live exams (Zaffar et al., 2023). Conventional algorithms have been replaced by deep-learning algorithms in object-detection research (Mittal et al., 2020). To use SVMs and AdaBoost for recognition, the features must be manually created in accordance with various experimental scenarios. In addition, these conventional object- detection algorithms have a more difficult feature-extraction method than deep-learning algorithms.

Deep learning was initially applied to object detection by introducing a R-CNN technique (Girshick et al., 2016). The Fast R-CNN method, an improvement over R-CNN,incorporates a spatial pyramid pooling network (SPPNet) (He et al., 2015), which greatly enhances the accuracy of object detection. However, both R-CNN and Fast R-CNN rely on selective-search algorithms, which are computationally expensive, memory-intensive, and slow to process when extracting regions. Despite the advancements made to the Faster R- CNN method (Ren et al., 2017), which increased the efficiency of candidate-area selection and employed anchor boxes of various scales, it still falls short in meeting real-time detection requirements and fails to address the problem of slow detection speed. This statement sounds confusing. Furthermore, the studies by Fang et al. (2021), Tan et al. (2021), and Deepa et al. (2019) introduced regression-based object detection techniques through their applications of SSD and YOLO algorithms. Xiao (2021) introduced the use of the SSD core detection algorithm based on deep learning for examination room monitoring, demonstrating its efficacy with an average accuracy of 79.8%, thereby improving the traditional monitoring methods. Potluri et al. (2023) introduced MG-Net, an advanced model using convolution operations and residual blocks for efficient exam proctoring with detecting multiple persons with an average accuracy of 75.3%. Despite slightly reduced detection accuracy, these methods achieve a detection speed of 23-24 frames per second (FPS), which is a substantial improvement over traditional methods that often process video data at lower frame rates, leading to delays in identifying suspicious activities. The higher FPS ensures real-time detection, enabling immediate responses to potential cheating behaviors during exams.

The most recent research has focused on enhancing the aforementioned algorithms. Zhang et al. (2019) a vehicle detection approach employing a modified faster region-based convolutional neural network (R-CNN). Li et al. (2022) proposed the depth-first search for detecting objects in remote-sensing photos. To improve the precision of object localization and enhance the detection accuracy for small objects, certain modifications were made to the dimensional clustering module, loss function, and detection approach using sliding- window segmentation. However, these modifications resulted in a noticeable decrease in the recall rate with 11.2%. Ting et al. (2019) enhanced the accuracy of the Faster R-CNN detection model by upgrading the training set and developing a two-channel network for feature extraction. However, the detection speed remained at 22 FPS, which, while marginally slower than the 23-24 FPS achieved by newer models such as MG-Net, is still considered insufficient for real-time detection in dynamic examination settings. The slight improvement in speed highlights progress in reducing computational overhead, but it falls short of meeting the requirements for immediate responses during live exam monitoring.

By examining the behavioral traits of item exchange, Lin et al. (2015) suggested a cheating-detection method based on a dynamic threshold. They utilized the iterative threshold method to establish a dynamic threshold for segmenting the differential images. The segmentation results were employed to guide background updates and facilitate the detection of cheating in the exam room through a background subtraction algorithm. However, they did not include test results, limiting the capability of the method to identify unusual item-exchange behavior. Dai et al. (2012) proposed a technique for detecting abnormal behavior in the exam room. The introduction of concepts such as behavior coverage area and 3D examination-room attention have brought new perspectives to the field. A model was developed using a latent SVM; however, the achieved accuracy with 72.4% and speed with 19FPS did not demonstrate clear superiority. Moreover, the detection range of this model is extremely limited owing to its practical application in real examination circumstances.

Several studies have attempted to apply YOLO-based models to cheating detection in examinations. Padhiyar et al. (2023) focused on developing an "Abnormal Behavior Detection Technique" for exams, addressing the inefficiencies of manual monitoring and emphasizing the importance of safety based on YOLO. This technique's strength is its focus on safety and the utilization of advanced object detection to monitor exams. Its weakness, however, is the potential for false positives in detecting abnormal behavior with accuracy of 73%. Ashwinkumar et al. (2021) developed an advanced algorithm leveragingtransfer learning and deep learning, integrating YOLO, MPGazeII, and VGG16 models, to enhance the reliability of examinations through automated proctoring and anomaly detection, overcoming the limitations of existing methods in terms of processing and feature extraction with accuracy of 79%. 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 with accuracy of 81%. Similarly, Alkhalisy et al. (2022) utilized a YOLOv5-based model to detect multiple cheating indicators, including the presence of additional persons in the examination area with accuracy of 81.5%. 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 negative, which can result in the misclassification of cheating instances. Additionally, their dataset was very limited, with only 3-6 classifications.

The rationale for using YOLO over other detection methods for cheating detection is multifaceted. Primarily, YOLO's architecture allows it to process images at a significantly higher speed than two-stage detectors like R-CNN, Fast R-CNN, and Faster R-CNN, which is critical for real-time applications (Redmon & Farhadi, 2018). YOLO's single- stage detection process, which encompasses simultaneous localization and classification, enables the algorithm to make immediate inferences, a necessity for catching cheating as it happens (Zaffar et al., 2023).

Furthermore, YOLO's successive iterations have shown continual improvements in dealing with the small objects and varied lighting conditions that are common in examination halls (Haque et al., 2023). The YOLOv3 iteration, for example, introduced multi-scale predictions and a more significant number of anchor boxes, improving the detection of small objects such as electronic devices or notes, which are often used in cheating (Redmon & Farhadi, 2018). Trabelsi et al. (2023) presented a deep learning-based cheating detection system using the YOLOv7 model, trained on a custom dataset, to enhance academic integrity by identifying cheating behaviors with a mAP@0.5 of 0.719, showcasing its potential in reducing errors in human monitoring.

The comparative effectiveness of YOLO is supported by empirical evidence showing its superior performance in object detection tasks. A study by Farhadi and Redmon (2018) demonstrated that YOLOv3 could process images at a speed of 22 milliseconds per image, significantly faster than the 106 milliseconds per image required by Faster R-CNN, while still maintaining a comparable level of accuracy of 79.12% on the COCO dataset. The importance of processing speed is underlined by the requirement for real-time cheating detection, where even minor delays can result in missed opportunities to catch cheating as it occurs, in general, 24 FPS is required.

Here's a summary table of the research reviewed for cheating detection:

Table 2.1: Summary table of the research reviewed for cheating detection

Study	**Algorithm Used**	**Strength**	**Weaknesses**	**Results Obtained**	**Data Used**
Xiao (2021)	SSD Core Detection Algorithm (Deep Learning)	High average accuracy	Reduced detection speed compared to newer methods	Average accuracy of 79.8%	Real-time video
Potluri et al. (2023)	MG-Net (Convolution Operations, Residual	Improved detection speed	Reduced average accuracy	Average accuracy of 75.3%, speed of	Real-time video

	Blocks)			23-24 FPS	
Li et al. (2022)	Depth-First Search, Dimensional Clustering, Sliding-Window Segmentation	Improved precision for small objects	Noticeable decrease in recall rate	Decreased recall rate by 11.2%	Static images (Remote-sensing photos)
Ting et al. (2019)	Faster R-CNN, Two-Channel Network for Feature Extraction	Enhanced detection accuracy	No improvement in detection speed	Accuracy improved, speed remained at 22 FPS	Real-time video
Dai et al. (2012)	Latent SVM, Behavior Coverage Area, 3D Attention Model	New perspectives on behavior detection	Low accuracy and speed	Accuracy of 72.4%, speed of 19 FPS	Real-time video
Alkhalis et al. (2022)	Deep Learning, Computer Vision	Focus on AI-based proctoring	Potential over-reliance on technology	Highlighted need during COVID-19	Real-time video
Padhiyar et al. (2023)	YOLO-based Abnormal Behavior Detection	Focus on safety	Potential for false positives	Accuracy of 73%	Real-time video
Ashwinkumar et al. (2021)	Transfer Learning (YOLO, MPGazeII,	Enhanced reliability through multiple	Complex integration of different models	Accuracy of 79%	Real-time video

	VGG16)	models			
Laurisa et al. (2022)	YOLOv4	Detection of unauthorized objects	May not handle all types of cheating effectively	Accuracy of 81%	Real-time video

Another significant advantage of YOLO is its robustness in varied environments. A robust detection system must perform well under different lighting conditions and various student densities within an exam hall. In the paper "YOLO9000: Better, Faster, Stronger," (Redmon et al., 2017) the authors addressed YOLO's improved robustness in diverse conditions, a feature that is especially pertinent to the dynamic and unpredictable environment of an examination hall (Redmon et al., 2016).

Moreover, the architecture of YOLO allows for customizations and improvements, such as the integration of attention mechanisms, which can be beneficial for cheating detection. Attention mechanisms help to focus the model on relevant features and areas within the video feed, thus enhancing the probability of detecting subtle cheating-related cues.

The application of YOLO in cheating detection is not without its challenges, such as the need for extensive datasets that include a range of cheating behaviors (Zaffar, 2023). However, the flexibility of the YOLO architecture enables the use of transfer learning,where a model pre-trained on a large dataset can be fine-tuned with a smaller, specific dataset related to cheating detection.

YOLO is distinguished by its relatively modest requirements for computational power, especially when compared to more resource-intensive models such as R-CNN

(Region- based Convolutional Neural Network) and its variants like Faster R-CNN and Mask R- CNN. These models are considered resource-intensive because they typically involve multiple stages in processing an image, including generating region proposals and then classifying each proposal using deep CNNs. This multi-stage processing demands significant computational resources, particularly in terms of memory and processing time (Zoev et al., 2019). In contrast, YOLO’s single-stage detection framework, which directly predicts bounding boxes and class probabilities from full images in one evaluation, allows it to operate efficiently even on hardware with limited capabilities. This efficiency makes YOLO an ideal choice for institutions that may lack specialized hardware, such as GPUs typically required by more complex models. Such practical consideration is crucial for scalable deployment across various examination settings where real-time processing is essential, and hardware capabilities might vary.

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 et al. 2023), can be utilized in applications such as object identification, image categorization, and instance segmentation. Ultralytics produced the influential YOLOv5 model, developed the YOLOv8. Compared with YOLOv5, YOLOv8 showed several architectural updates and enhancements.

2.6 Dataset

In reviewing the existing literature and conducting extensive research, it has become increasingly apparent that the datasets employed in the development of object detection algorithms for exam cheating detection are predominantly proprietary. These datasets are specifically designed for individual models, leading to a significant gap in publicly accessible resources. Presently, there is a conspicuous lack of datasets that are both high-resolution and rich in labelled data, which severely limits their effectiveness and applicability in broader contexts. This scarcity highlights the critical need for the development of an open-source, high-definition dataset that is not only comprehensive in terms of labeled instances but also validated for its effectiveness in real-world scenarios. The creation of such a dataset would represent a significant advancement in the field, providing a valuable resource for future research and development. In recognition of this need, this research has undertaken the initiative to develop such a dataset. The specifics of this dataset, including its structure, labeling methodology, and potential applications, will be thoroughly discussed in Section 3.1. Table 2.2 shows the number of labels and the size of datasets in the literature review, as well as the corresponding results. It can be seen that the data set constructed in this paper is the most comprehensive so far, with the largest number of data sets and the most reasonable number of labels.

Table 2.2: Dataset size and label in literature review

Name	Size of dataset	Number of labels	Results
Zaffar et al. (2023)	9056	6 types NoCheat; LookLeft; LookRight; PocketSheet; PantsSheet; ExchPaper;	81.67%
Xiao (2021)	10927	4 types Mobile phone/copy; Peeking at the answers; Passing the answers	79.8%
Potluri et al. (2023)	2041	2 types Normal; Abnormal;	75.3%
Ashwinkumar et al. (2021)	1000	2 types Normal; Abnormal;	79%
Laurisa et al. (2022)	1827	3 types Normal; Use mobile phones; Use Books	81%
Alkhalisy et al. (2022)	6300	5 types	81.5%

		App Tasks; Mobile Device; Head Pose; Multi person; Mouth Open	

2.7 Summary

The literature review has provided a critical examination of the existing research and developments that form the foundation of this thesis. Computer vision, serving as the bedrock of automated visual tasks, has been discussed with a focus on its application in monitoring and detecting anomalies in examination settings. We delved into CNNs, the class of deep neural networks most adept at processing visual imagery, laying out their architectural principles and the advancements that have propelled their success in object recognition. A subsequent exploration of object detection algorithms, such as R- CNNs, YOLO, and SSDs, provided insight into their operational mechanisms and comparative performances.

In the realm of cheating detection, we scrutinized the application of these algorithms, with particular attention paid to the use of YOLOv8 variants and the integration of attention mechanisms to refine detection capabilities. The efficacy of these models in discerning cheating behavior from legitimate academic conduct was discussed, underscoring the technological strides made in this area. The datasets section emphasized the critical role of high-quality, annotated visual data in training accurate and reliable detection systems. We highlighted the challenges of dataset creation, such as the need for diversity in behavior representation and the extensive labor required for annotation, while also acknowledging the value of such datasets in pushing the boundaries of what is achievable with current AI technologies.

Through this review, it is evident that the field is on the cusp of significant breakthroughs, with the potential for technologies like YOLOv8 integrated with attention mechanisms to revolutionize cheating detection. Yet, there remains a clear need for ongoing research to address limitations such as the context understanding and real-time processing demands in diverse examination environments. The insights garnered here set the groundwork for the novel contributions of this research, which aims to refine these technologies further and expand their applicability, ensuring fair and credible examination conditions. The ensuing chapters build upon this foundation, presenting the methodological innovations and empirical validations that constitute the core contributions of this thesis.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The methodology employed in a research project is the blueprint for achieving its objectives, providing a systematic approach that underpins the study's validity. This chapter begins by detailing the data collection process, which forms the foundation of the study by providing the necessary inputs for training and evaluating the proposed model. Subsequently, the YOLO algorithms, central to our study's object detection strategy, are discussed in depth. The chapter explores the evolution of YOLO algorithms, their operational paradigms, and their suitability for real-time object detection tasks. Following this, attention mechanisms, a relatively recent innovation in deep learning, are introduced, elucidating how they augment the YOLO algorithms' capabilities by enhancing feature selectivity and focus. Building on this foundation, we present our methodological contribution: the improvement of the YOLO algorithm with a specific type of attention mechanism, the Efficient Channel Attention, tailored for the detection of cheating behaviors in examination settings. This integration aims to address the challenges posed by conventional surveillance methods in accurately identifying cheating incidents. By employing a combination of theoretical insights and empirical experimentation, this chapter delineates the step-by-step process of the proposed method, from conceptualization to computational implementation, ultimately leading to the validation of the enhanced YOLOv8 model's effectiveness in a controlled testing environment. The complete methodology is illustrated in Figure 3.1. First collect data, then execute and implement improvements and optimizations of YOLOv8.

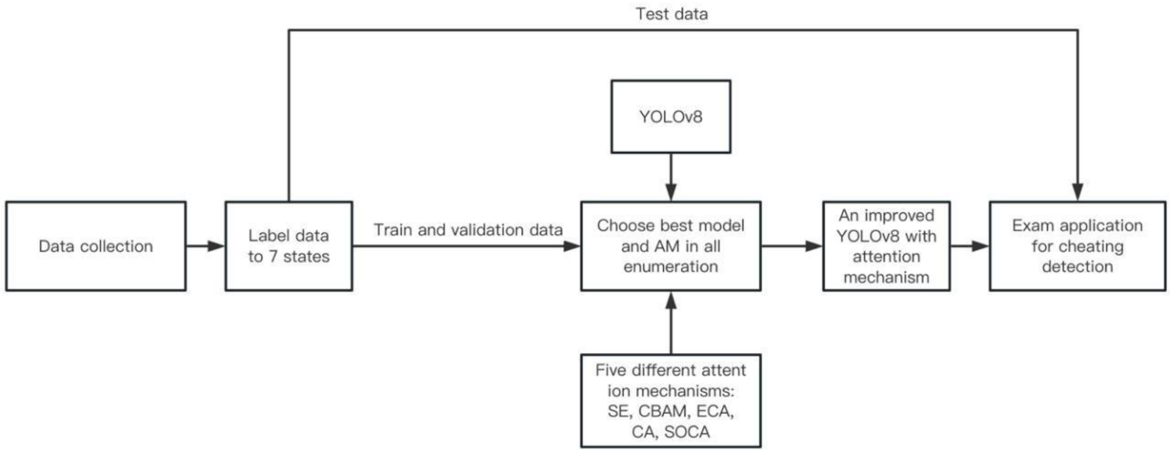

Figure 3.1: Methodology of the study

3.2 Data collection

To overcome the scarcity of publicly available datasets that specifically focused on students' anomalous behaviors during exams, particularly cheating, we designed and prepared our own dataset. This dataset is specifically curated to encompass a wide range of actions that students may engage in during paper-based examinations, including behaviors that could potentially facilitate cheating. It covers the majority of cheating techniques as well as normal behaviors, including normal, passing items (notes, rulers, etc.), whispering, placing hands under the table, taking out mobile phones, walking around, and looking around. Figure 3.2 shows a visual representation of the distinct states captured in the dataset.

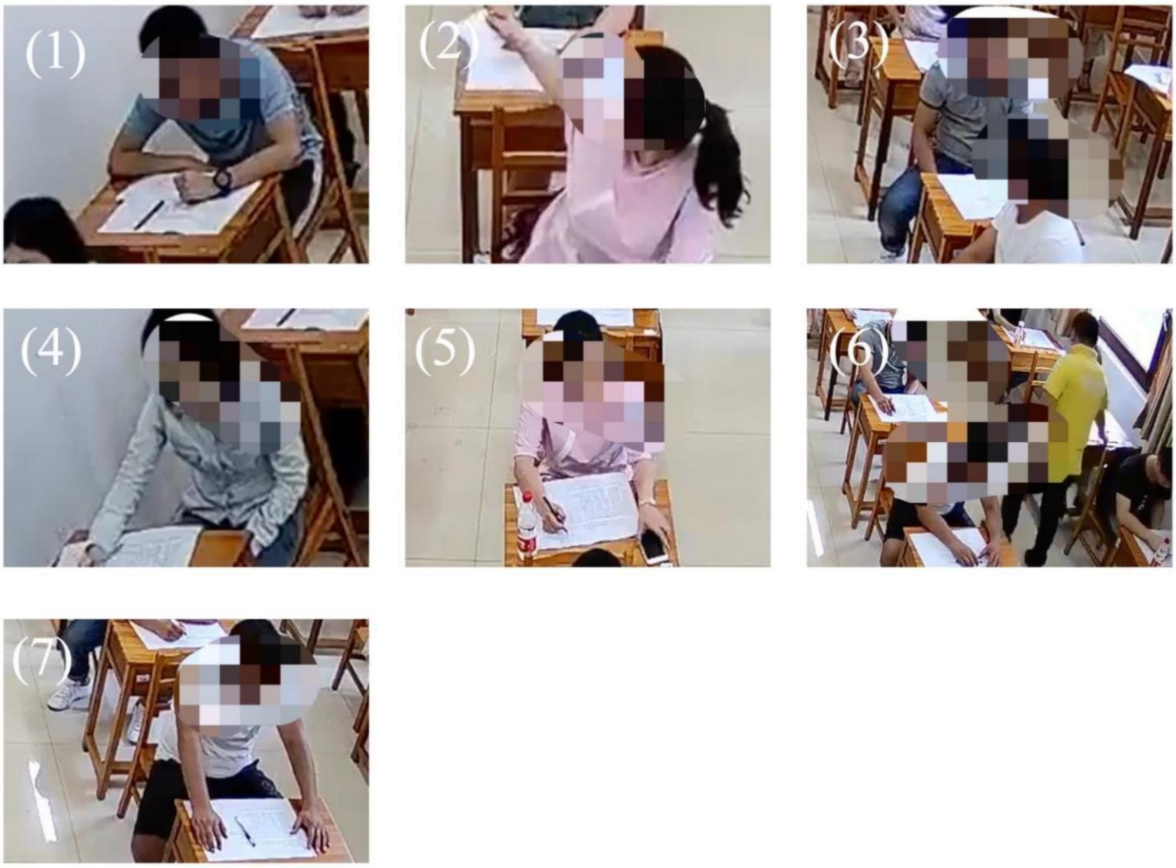

Figure 3.2: Seven different states: 1. normal; 2. passing items (notes, rulers, etc.); 3. whispering; 4. placing hands under the table; 5. taking out mobile phones;6. walking around and 7. looking around captured in the dataset

The scenes were captured using a Sony A7M4 camera with a Sony 30–170 mm lens in a high school classroom in China with 26 students. The camera recorded at 30 frames per second with a resolution of 1920 × 1080 pixels. This frame rate is suitable for detecting actions, even subtle movements, without missing important details. The camera, placed at the top of the classroom, focused on capturing the hand regions of the subjects. Multiple recording sessions were conducted, which were scripted and acted to simulate realistic examination scenarios. These recordings were not taken during actual exam periods to ensure a controlled environment where specific behaviors could be captured systematically. The scenarios were designed to include both typical student behaviors and potentially anomalous actions indicative of cheating.

All participants were fully informed about the recording process and provided written consent prior to the sessions. The purpose of the study and the nature of the simulated scenarios were explained to them to ensure ethical compliance. To maintain consistency, protocols were established and adhered to during each recording session. The camera settings were standardized, and the classroom layout was kept consistent across all datasets. Additionally, the duration of each recording session was set to 20 mins, matching the average length of an exam in a high school setting, thereby providing realistic and contextually relevant data for analysis.

The proposed dataset introduces inherent challenges and complexities due to the similarity between various activities and the inability to solely determine actions based on body movements. For example, behaviors such as “drinking water” or “resting on the desk” should be regarded as normal, requiring additional contextual information to distinguish them from potentially anomalous actions. Therefore, when analyzing the dataset, factors such as the subject’s proximity to objects and their underlying intentions must be considered. By incorporating these elements, the study aims to achieve a more accurate detection and classification of anomalous actions, enhancing the system’s robustness in real-world scenarios.

3.3 YOLO algorithms

YOLO is an object-detection algorithm that is a type of computer-vision technique aimed at identifying and localizing objects within images or video frames. The primary goal of object-detection algorithms is to recognize objects belonging to various predefined classes and determine their locations within an image using bounding boxes. The state-of- the-art YOLOv8 is a good option for various object-recognition and image-segmentation tasks because it is quick, precise, and simple to use. It can run on a range of hardware platforms, from CPUs to GPUs, and can be trained on large datasets. The YOLOv8 architecture is shown in Figure 3.3. The architecture of YOLOv8 presented in the diagram illustrates a sophisticated deep learning model for object detection in real-time. It is comprised of two primary components: the Backbone and the Head.

The Backbone serves as the feature extractor, where convolutions are employed at various stages (P1 to P5) to process the input image and extract features across multiple scales. It is built upon the CSP network structure, integrating several elements like diverse convolutional layers with different kernel sizes for downsampling, C2F modules for efficient feature utilization and computation reduction, as well as the SPPF block to amalgamate features from various scales.

On the other hand, the Head, also known as YOLOv8Head, is tasked with object detection using the features garnered by the Backbone. It is designed to detect objects at multiple scales through several detection layers, employing a combination of convolutions, up-sampling, and concatenation of feature maps. These feature maps from the Backbone are enriched with semantic information which is crucial for accurate detection. YOLO's grid-based detection approach is applied here, entailing bounding box regression and class prediction.

For the training process, a mixture of loss functions is utilized, often including Binary Cross-Entropy for class prediction accuracy, and a form of regression loss for bounding box precision.

In summary, the YOLOv8 architecture integrates deep convolutional networks with a multi-scale detection mechanism, optimizing the process for fast and efficient real-time object detection.

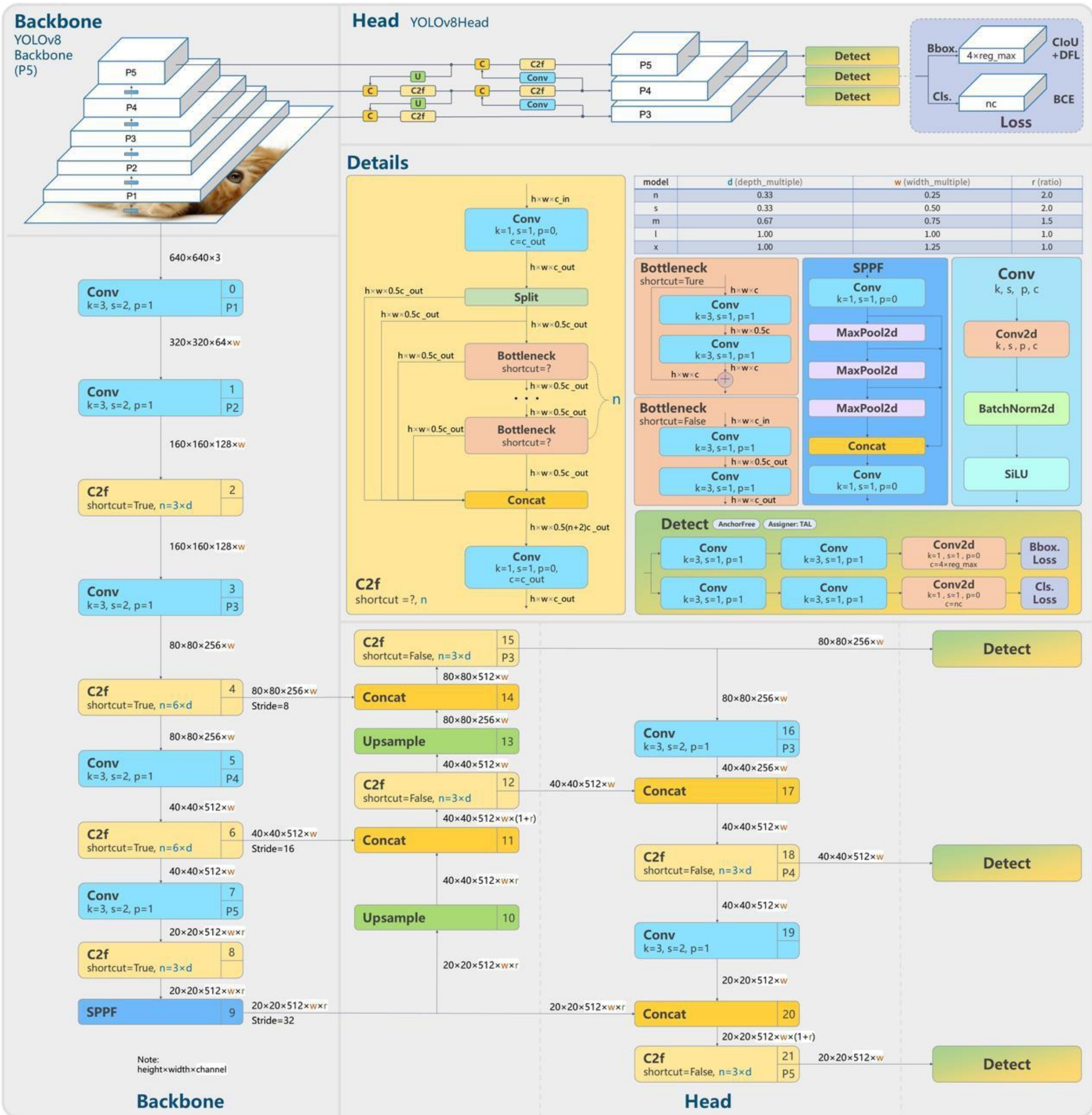

Figure 3.3: YOLOv8 architecture (Terven et al., 2023)

YOLOv8 utilizes Darknet as its underlying architecture, akin to the role of ResNet in other models. Darknet is a lightweight neural network framework designed for object detection tasks, emphasizing speed and efficiency without compromising performance. It supports the construction of deep CNNs and is optimized for high-speed inference.

Network degradation in CNNs refers to the phenomenon where, paradoxically, the network's performance starts to deteriorate as it becomes deeper, instead of improving. This issue arises due to difficulties in training very deep networks, including problems with gradient flow that can lead to vanishing or exploding gradients. To overcome this, the concept of a residual block was introduced, enabling the construction of deeper networks by facilitating the efficient transmission of parameters between layers. This innovation ensures that deeper networks can learn effectively without succumbing to degradation, thereby enhancing their learning capacity and accuracy.

When Darknet is employed for object detection, the fully connected (FC) layer is removed and convolutional layers are utilized instead. Figure 3.4 presents a comprehensive visualization of the YOLOv8 network architecture.

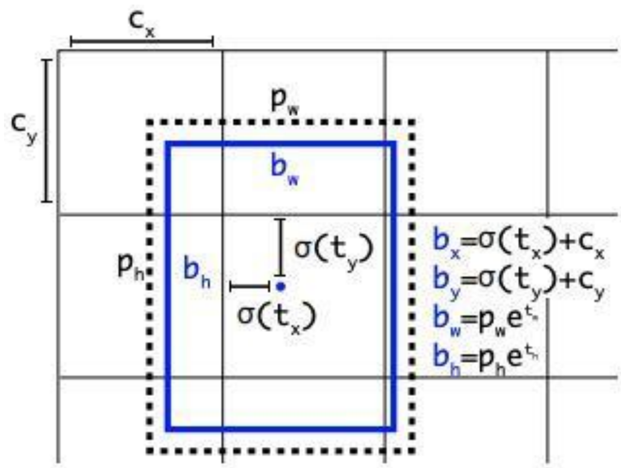

Figure 3.4: Visualization of an anchor box in YOLO (Solawetz, 2020)

In the latest iteration of YOLOv8, several modifications were implemented to boost its performance. To begin with, the C3 module, a core component of YOLO's architecture known for combining convolutional and cross-stage partial layers to enhance feature extraction, was replaced. Its successor, the C2f module, introduces an optimized design aimed at refining the model's ability to process and interpret image data more efficiently, though the specific function of the C2f module extends beyond traditional feature extraction to possibly include more advanced integration of features across different scales.

Furthermore, the initial 6×6 convolutional layer found in the backbone of the model, which is crucial for the initial stages of feature extraction, was substituted with a more standard 3×3 convolutional layer. This change is significant because 3×3 kernels are widely recognized for their effectiveness in capturing the essential details and spatial hierarchies in images with less computational overhead compared to larger kernels.

Additionally, modifications were made within the bottleneck structure of the model, where the first 1×1 convolutional layer was replaced. Previously, 1×1 convolutions were employed to reduce the dimensionality of the feature maps, thereby decreasing the computational load while preserving important information. This layer was substituted with a 3×3 convolutional layer, a decision likely aimed at enhancing the model's feature extraction capabilities without significantly increasing the computational burden. This adjustment suggests a strategic trade-off, prioritizing the extraction of richer feature representations over the model's compactness and efficiency.

Furthermore, a decoupled head was employed and the objectness branch was removed. Notably, YOLOv8 is an anchor-free model, that is, it directly predicts the center of an object instead of estimating the offset from a predefined anchor box.

In this study, anchor-free detection was utilized to reduce the number of box predictions, thereby speeding up the nonmaximum suppression process, which is a complex post-processing step that involves sorting through candidate detections after inference (Figure 3.3).

3.4 Attention Mechanism

In computer vision (CV), attention mechanisms are inspired by how humans selectively focus on specific parts of a scene while processing visual information (Hafiz et al., 2021). These mechanisms enable neural network 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). 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 each channel, this mechanism employs an excitation operation comprising 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 (Woo et al., 2018), CBAM can extract feature information that demands thorough attention. These attention mechanisms will be used to integrate with YOLOv8, enhancing the detection performance of cheating behavior.

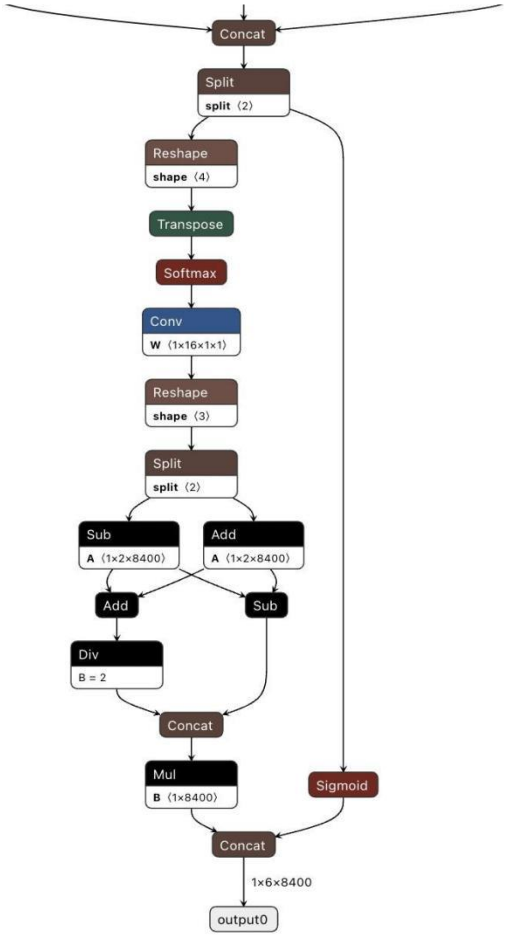

Figure 3.5: Detection head for YOLOv8

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 (Xu et al., 2018). A prominent example is the SENet which employs attention mechanisms to dynamically recalibrate feature responses on a channel-wise basis. Attention mechanisms have been widely applied in object-detection tasks. 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 an improved object-detection accuracy (Lin et al., 2020). In Yu et al., (2022), an SE network that explicitly modelled the interdependencies between channels in CNNs was proposed. Another study (Hu et al., 2020) 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 (e.g., image classification, object detection, and image segmentation). These studies collectively illustrate the transformative impact of attention mechanisms, heralding a shift in the paradigm of visual understanding and recognition.

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 (Hafiz et al., 2021). 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 (Jin et al., 2021). This resilience to occlusions further enhances the performance and robustness of the models. Figure 3.6 illustrates the simplified network topology of YOLOv8, in which the backbone extracts the input image data, and the feature fuse combines them to obtain more comprehensive target features for accurate predictions. Figure 3.7 shows an overview of several typical attention methods.

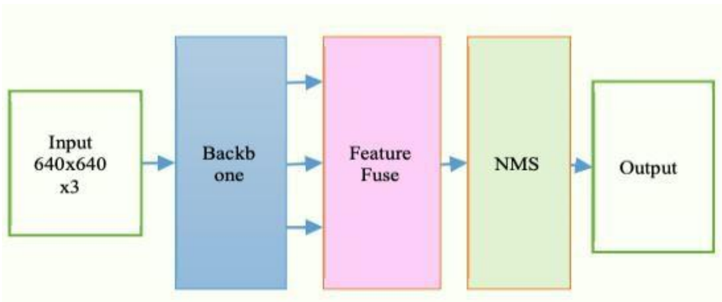

Figure 3.6: Main flowchart of Yolov8

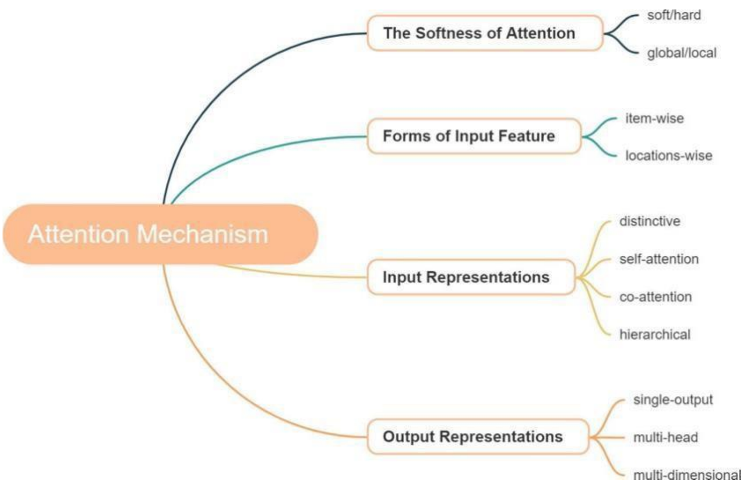

Figure 3.7: Overview of several typical attention methods

3.5 Improvement of YOLO with Attention Mechanism

Although it is one of the most advanced versions of the YOLO architecture, YOLOv8 still possesses certain limitations. For example, it displays a reduced capacity to capture the rich contextual information inherent within the images (Pathak et al., 2016). This is because YOLOv8 follows a grid-based approach, wherein each grid cell independently predicts the bounding boxes and object classes. Such an approach can result in underutilization of contextual relationships between different parts of an image (Vercauteren et al., 2019). The addition of an attention mechanism effectively addressed these limitations (Sourial et al., 2021). By modelling object-context relationships explicitly and selectively, focusing on informative cues within the image, attention mechanisms incorporated into this model could help to improve the model's ability to detect and classify objects. This is achieved by generating an attention map that highlights regions in the image that are deemed critical for accurate detection, thus reducing the chances of missing important contextual information. The addition of an attention mechanism to YOLOv8 can potentially improve its performance by allowing the model to focus on the more relevant regions of the input image. Furthermore, attention mechanisms have been shown to improve the performance of various deep-learning models, especially in CV tasks (Wang et al., 2022, Yu et al., 2022).

However, although the aforementioned studies demonstrated the potential benefits of integrating attention mechanisms with object-detection models, most prior research has focused on testing a single type of attention mechanism. This study addresses this limitation by evaluating multiple attention mechanisms within the YOLOv8 architecture, providing a more comprehensive analysis of their effects on detection performance.

Cheating detection in an educational environment presents a unique set of challenges. Cheating activities often involve subtle behaviors or objects that conventional object detectors can overlook. In contrast, attention mechanisms integrated with YOLOv8 can significantly enhance cheating detection. By explicitly modelling object–context relationships and focusing on relevant regions, attention-augmented YOLOv8 can detect subtle cheating indicators more accurately.

Thus, the incorporation of attention mechanisms with YOLOv8 is a promising solution for enhancing the robustness and accuracy of cheating detection. By addressing the limitations of YOLOv8 and drawing upon the strengths of the attention mechanisms, this combination might significantly boost exam integrity. The improved YOLOv8 algorithm, which uses CNNs as its backbone, benefits from such depth and complexity. The system's effectiveness in detecting cheating behaviors is primarily dependent on how well the CNN is trained on relevant data. This training includes not only images that capture the act of cheating but also a broad range of normal behaviors to which the model must learn to respond with high specificity and sensitivity. The architecture's ability to generalize from the training data to real-world scenarios is critical for a successful deployment in examination halls.

The attention mechanism assigns weights to different parts of the input data, signifying their relevance to the current task. In CV, weights are assigned to the pixels or regions in the input image. Although YOLOv8 is a leading object-detection model, its performance in detecting small or subtle cheating signs, which may occupy a small portion of the image or face distractions, can be further enhanced by integrating the selected attention mechanism into the model architecture, specifically within the feature-extraction phase or backbone (Ma et al., 2021). 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 (Hu et al., 2020), Convolutional Block Attention Module (CBAM) (Woo et al., 2018), efficient channel attention (ECA) (Wang et al., 2020), coordinate attention (CA) (Hou et al., 2021), and second-order attention network (SOCA) (Dai et al., 2019).

Table 3.1: Summary of attention models.

Citation	Model	Key Characteristics
Hu et al., 2020	Squeeze-and-Excitation Networks (SENets)	Introduces the Squeeze-and-Excitation (SE) block which adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels.
Woo et al., 2018	Convolutional Block Attention Module (CBAM)	Combines both channel and spatial attention mechanisms to refine features by considering 'what' and 'where' is important. Integrates channel attention and spatial attention sequentially.
Wang et al., 2020	Efficient Channel Attention (ECA)	Proposes a lightweight channel attention mechanism without dimensionality reduction. Uses a 1D convolution with adaptive kernel sizes to capture local cross-channel interactions efficiently.

Hou et al., 2021	Coordinate Attention (CA)	Encodes both channel and spatial information into channel attention. Captures long-range dependencies along one spatial direction while preserving precise positional information along another.
Dai et al., 2019	Second-order Attention Network (SOCA)	Utilizes second-order statistics to enhance the feature representation. Employs covariance pooling to capture richer and more discriminative information for attention mechanisms.

In this study, we used the ECA mechanism that show the best detection performance for cheating in later Chapter 4. This mechanism specifically addresses the limitations of the existing channel attention methods, which often involve complex operations and high computational costs (Chen et al., 2023). Inspired by the concept of local cross-channel interaction, the ECA mechanism recognizes that spatially neighboring channels share similar semantic information. To capture these local interactions efficiently, the ECA mechanism employs a lightweight one-dimensional convolution network, which reduces the computational complexity typically associated with conventional channel attention methods. The ECA module utilizes a kernel size k, which is adaptively determined based on the number of channels in the input feature map. This adaptability is achieved through a mapping function that relates the kernel size k to the number of input channels C. Instead of using a fixed kernel size across all scenarios, the ECA mechanism dynamically computes k according to the equation $k = \Psi(C)$, where Ψ is a non-linear mapping function. This ensures that the kernel size is neither too small, which might fail to capture sufficient local dependencies, nor too large, which could introduce redundant computations. By adapting the kernel size to the channel dimensions, the ECA mechanism achieves a balance between efficiency and effectiveness, enabling it to capture meaningful local dependencies across varying input sizes while maintaining a compact model size. Figure 3.8 shows the structure of the ECA module. Given the aggregated features obtained by global average pooling (GAP), ECA generates channel weights by performing a fast 1D convolution of size k, where k is adaptively determined by mapping the channel dimension, C. (Xiang et al., 2021)

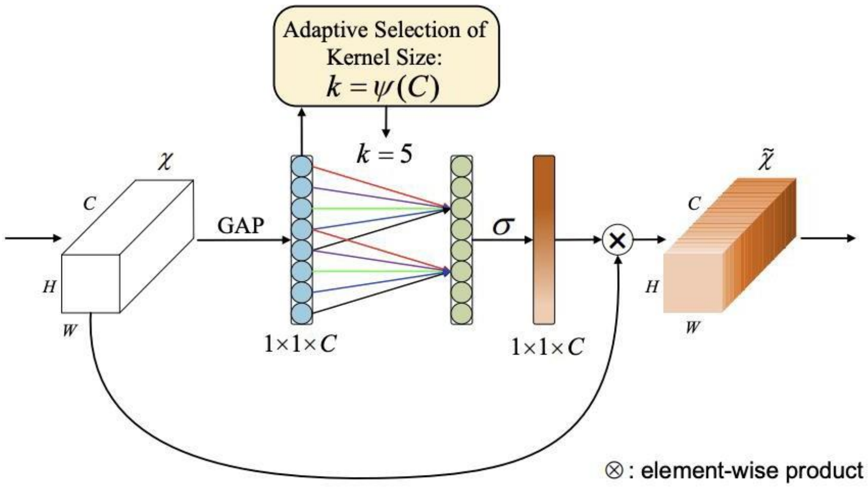

Figure 3.8: Diagram of the ECA module

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 ensure the integrity of the network's inherent characteristics, we made a strategic decision to incorporate an attention mechanism within the last feature-extraction layer.

To achieve this, we introduced an attention layer in the ECA module, following the first convolutional layers within the backbone network, as depicted in Figure 3.9.

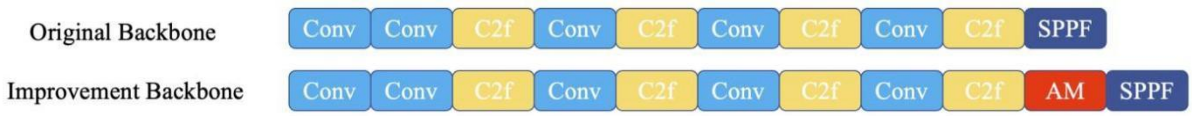

Figure 3.9: Improvement for YOLOv8 by using an attention mechanism (AM)

This placement allowed us to effectively integrate the attention mechanism while preserving the vital properties of the extracted network. By strategically inserting the ECA module at this stage, we can maintain the network’s functionality and prevent any negative impact on its prediction capabilities.

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. The overall flow is summarized in the provided pseudocode.

The initial forward pass is made through the selected YOLOv8 architecture's backbone, producing a set of feature maps:

Table 3.2: Algorithm1: YOLOv8 and attention modules Initialization

Algorithm1: YOLOv8 and attention modules Initialization

```
Initialize YOLOv8 Network using the Darknet backbone for feature extraction.
Initialize SE, CBAM, ECA, CA, and SOCA attention modules

for epoch in 1...N:
    for batch in DataLoader:
        images, ground_truths = batch

        // Forward pass through initial YOLOv8 layers
        feature_maps = YOLOv8_Backbone(images)
```

Then, we introduced a novel architecture that integrates multiple attention mechanisms, namely SE, CBAM, ECA, CA, and SOCA, with the YOLOv8 object detection framework. As shown in Algorithm 2.

Table 3.3: Algorithm2: A novel integration of multiple attention mechanisms with the YOLOv8 object detection framework

Algorithm2: A novel integration of multiple attention mechanisms with the YOLOv8 object detection framework

```
// Squeeze-and-Excitation (SE) Attention
for fm in feature_maps:
    SE_out = SE_Block(fm)
    fm = fm * SE_out

// Convolutional Block Attention Module (CBAM) Attention
for fm in feature_maps:
    CBAM_out = CBAM_Block(fm)
    fm = fm * CBAM_out

// Efficient Channel Attention (ECA) Attention
for fm in feature_maps:
    ECA_out = ECA_Block(fm)
    fm = fm * ECA_out

// Coordinate Attention (CA) Attention
for fm in feature_maps:
    CA_out = CA_Block(fm)
    fm = fm * CA_out

// Second-order Attention Network (SOCA) Attention
for fm in feature_maps:
    SOCA_out = SOCA_Block(fm)
    fm = fm * SOCA_out
```

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.

The YOLOv8 loss function consists of three main components:

1. Localization Loss: This component evaluates the errors in the predicted bounding box coordinates, ensuring the detected objects are accurately localized within the image.

2. Objectness Loss: This measures the errors in the predicted objectness scores, determining the likelihood of an object being present in each bounding box.

3. Classification Loss: This assesses the errors in the class probability predictions, ensuring accurate classification of the detected objects.

Table 3.4: Algorithm3: Object Detection and Classification Algorithm

Algorithm3: Object Detection and Classification

```
    // Continue with standard YOLOv8 object detection and classification
    detections = YOLOv8 _Head(feature_maps)

    // Calculate Loss
    loc_loss = MSE(detections.bboxes, ground_truths.bboxes)
    obj_loss = MSE(detections.obj_scores, ground_truths.obj_labels)
    class_loss = CrossEntropy(detections.class_probs, ground_truths.class_labels)
    total_loss = lambda_loc * loc_loss + lambda_obj * obj_loss + lambda_class *
class_loss
```

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 (Kingma, 2014).

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.

Table 3.5: Algorithm 4:Training and Validation

```
Algorithm4:Training and Validation

        // Backward pass
        Backpropagate(total_loss)

        // Update weights
        UpdateWeights()

    // End of one epoch
Validate_on_ValidationSet()
```

3.6 Summary

To summarize, this chapter has presented a comprehensive overview of the methodologies applied in the development of an improved cheating detection system. The discussion commenced with a deep dive into the YOLO algorithms, which form the foundation of our object detection approach. The series of YOLO models, from their inception to the latest iteration used in this study, were explored to provide a contextual backdrop for their selection and use in our research. The attention mechanisms section illuminated the transformative impact these mechanisms have on the performance of deep learning models, particularly in tasks requiring discerning focus, such as the detection of nuanced cheating behaviors.

The innovative core of methodology section was the detailed exposition of the improvements made to the YOLOv8 algorithm through the integration of the ECA mechanism. By customizing the YOLO architecture with ECA, we addressed the specific demands of cheating detection in examination venues—a novel approach in the field of academic integrity. This integration is expected to yield a model that is not only precise in distinguishing between cheating and non-cheating actions but also robust across various exam conditions and student behaviors. The experimental setup and execution were also discussed, outlining the procedures for training, validating, and testing the model using the datasets described in the previous chapter.

CHAPTER 4

RESULTS AND ANALYSIS

4.1 Introduction

The integrity of empirical research is contingent upon the robustness of its results and the subsequent analysis thereof. This chapter critically examines the results obtained from the implementation of an improved YOLOv8 algorithm with an attention mechanism for detecting cheating in examination settings. This section sets the stage for a thorough discussion on dataset preparation, experiment setup, performance evaluation, and an in-depth analysis of the results garnered from the computational experiments conducted as part of this study.

An extensive dataset, specifically curated to include a range of cheating behaviors within a simulated examination environment, forms the basis of the evaluation. The setup for the experiment is described, detailing the configuration of the YOLOv8 algorithm and the integration of various attention mechanisms aimed at enhancing the model's detection capabilities. Performance results are then presented, providing quantitative data on the accuracy, speed, and reliability of the detection system across different cheating scenarios.

The analysis is expanded to interpret these results in the context of the system's practical application, considering the complexity of the examination process and the subtleties involved in defining and detecting cheating. This section not only addresses the performance of the algorithm in a controlled environment but also extrapolates its effectiveness to real-world settings, where variables and conditions are more dynamic and unpredictable. The introduction serves as a precursor to a comprehensive evaluation of the study's findings, highlighting the novel contributions of this research and setting the scene for the critical analysis that follows.

4.2 Dataset Description and Processing

The dataset is a fundamental component for the training and validation of any machine learning model, and its quality, as much as its quantity, plays a crucial role in the potential success of the algorithm in practical applications. In the case of study, the dataset for cheating detection during examinations was meticulously curated through the simulation of examination environments via video recording, reflecting the high-stakes conditions under which the model is expected to operate.

For the purposes of this research, we generated two 20-minute high-definition videos, simulating an examination setting with multiple subjects engaging in various activities, some of which were instances of cheating, while others constituted normal, permissible behavior during an exam. The total footage amounted to 40 minutes, which, when broken down into frames, produced a dataset of 72,000 images. This substantial volume of visual data offers a comprehensive basis for the training, validation, and testing of the improved YOLOv8 model.

Labeling such an extensive dataset is a task of herculean proportions, requiring in excess of a million annotations to accurately mark cheating behaviors across the collected images. The labeling process involved classifying each image with one or more of seven distinct label types, reflective of the different cheating methods and objects to be detected. These labels covered a range of cheating activities, from the use of unauthorized notes to signaling between students, as well as the use of electronic devices. The distribution of these labels varied, with some categories having a significantly higher representation in the dataset than others. The label counts for each category are shown in Table 4.1.

Table 4.1: Numbers of samples for each label

Label	Number	
	Train and validation (Video 1)	Testing (Video 2)
Normal	2593	1385
Passing Items	1126	673
Whispering	1456	697
Placing Hands Under The Table	1124	539
Taking Out Mobile Phones	523	264
Walking Around	556	295
Looking Around	569	261

Given the impracticality of labeling every frame due to time and resource constraints, we adopted a strategy of selective annotation. From the first video, 500 frames were manually selected to ensure that they included a diverse range of labels representing all seven defined behaviors. This selection process was guided by human observation to maximize the representation of the different actions, such as normal behavior, passing items, whispering, placing hands under the table, taking out mobile phones, walking around, and looking around. By focusing on frames with varied and meaningful behaviors, this approach ensured that the annotated dataset captured the complexity and diversity of cheating behaviors, providing a robust foundation for model training and evaluation. This approach is grounded in the understanding that, within a continuous video sequence, not every frame is essential for capturing the contextual nuances of an action. By focusing on key frames, which encapsulate significant movements or changes in the scene, we ensure that the model is trained on the most informative data points that adequately represent each category of cheating behavior.

To prepare the dataset for model training, we partitioned the labelled images into training and validation sets, allocating 80% for training and 20% for validation within each category. This split was designed to provide the model with a substantial learning base (the training set), while reserving a subset of data to evaluate the model's generalization capabilities on unseen data (the validation set). The validation set's role is critical as it serves as a benchmark for tuning the model's hyperparameters and for early detection of overfitting, where the model might learn to recognize patterns specific to the training set that do not generalize well to new data.

The second video was entirely withheld from the training process and used as a test set to evaluate the model's performance under conditions that it had not previously 'seen'. The analogous distribution of labels in the test set (outlined in Table 4.1) ensures that the model's accuracy and robustness are tested against a balanced array of instances across the spectrum of cheating behaviors.

To assess the model's performance, we adopted a dual approach, analyzing both the loss of the training and validation sets and the prediction accuracy on the validation set. The accuracy is a measure of the frequency with which the model incorrectly labels a frame – either through false positives, where a non-cheating action is labelled as cheating, or false negatives, where a cheating action goes undetected. This metric is critical for understanding the model's precision and recall, two performance indicators that are especially pertinent in the context of cheating detection, where both types of errors carry significant consequences.

Prediction accuracy, on the other hand, provides a direct measure of the model's competency in correctly identifying and classifying frames according to the predefined labels. It is calculated as the proportion of true positive detections against the total number of frames where a cheating action is present. In the validation phase, high prediction accuracy signifies a model well-tuned to the specificities of cheating behaviors, capable of discerning between legitimate and dishonest actions with high reliability.

In processing the dataset, we applied standard image preprocessing techniques such as resizing, normalization, and data augmentation to ensure uniformity across the dataset and to enrich the model's exposure to varied representations of cheating behavior. Data augmentation techniques, such as random cropping, rotation, and flipping, were employed to simulate different perspectives and angles, thereby enhancing the model's robustness and its ability to generalize from the training data to real-world testing scenarios.

4.3 Experiment 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 labelling 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 labelled data with YOLOv8, a script was employed to convert the JSON files into a TXT format that can be read and utilized for training.

In January 2023, five models of YOLOv8 were launched, as demonstrated in Table 4.2. These models provide the following key information:

Size (pixels): This parameter represents the input resolution of the image.

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 to 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 exchange (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.

Table 4.2: Characteristics of the five YOLOv8 models.(Jocher et al., 2023)

Model	Size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	Params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

As the dataset utilized in this study is unique, a comprehensive debugging and training process was conducted to evaluate all five attention mechanisms SE, CBAM, ECA, CA, and SOCA 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 data-augmentation methods applied in this study included random cropping, where the crop size was randomly selected within a range of 0.7 to 1.0 of the original image size while maintaining the aspect ratio. Random resizing and scaling were performed, resizing images to scales ranging from 0.5x to 2.0x of their original dimensions. Horizontal flipping was applied with a probability of 0.5, and random rotations were applied with angles randomly selected within ±10 degrees. Brightness adjustments ranged within ±20%, while contrast was varied by ±15%, and hue adjustments ranged within ±10 degrees with saturation altered by ±15%. These augmentations introduced variability into the training data, helping the model generalize better to unseen data. Furthermore, the parameter settings of the YOLOv8 models, including learning rate, batch size, and optimizer type, were preserved at their default values. The training process was conducted over 1000 epochs to ensure sufficient learning from the augmented dataset.

4.4 Performance Results

The analysis of the data provided in Table 4.3 reveals a comprehensive evaluation of different configurations of the YOLOv8 model integrated with various attention mechanisms. The table presents a comparison across multiple iterations of YOLOv8, including the standard model (denoted as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x) and those appended with attention mechanisms, such as SE, CBAM, ECA, CA, and SOCA. These configurations are evaluated based on their training and validation losses across three categories: box_loss, obj_loss, and cls_loss.

Table 4.3: Training results for different combination of YOLOv8 variants with different models of attention mechanism

Item	train	train	train	val	val	val
	box_loss	obj_loss	cls_loss	box_loss	obj_loss	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	0.00720
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	0.02010	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	0.00200	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	0.00680	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

Box loss measures the accuracy of the predicted bounding boxes' coordinates compared to the ground truth boxes. It ensures that the predicted boxes closely match the actual locations and dimensions of the objects in the image. Box loss typically includes components such as: Measures the difference in the coordinates (x, y) of the bounding box center. Measures the difference in the dimensions (width, height) of the bounding box. The formula for box loss can be expressed as:

$$\text{box}_{loss} = \sum_{i=1}^{N} \left(\text{smooth}_{L1}(\hat{x}_i - x_i) + \text{smooth}_{L1}(\hat{y}_i - y_i) + \text{smooth}_{L1}(\hat{w}_i - w_i) + \text{smooth}_{L1}(\hat{h}_i - h_i)\right) \quad \text{Equation 4.1}$$

SmoothL1 is a loss function that is less sensitive to outliers compared to L2 loss. L2 loss, also known as Mean Squared Error (MSE), measures the squared differences between the predicted and true values. While L2 loss is effective for penalizing large errors due to its quadratic nature, it can overly emphasize outliers, making the model less robust. In contrast, SmoothL1 loss combines the advantages of L1 and L2 loss by behaving like L2 loss for small errors and like L1 loss for large errors, thereby reducing the impact of outliers. Accurate box predictions are crucial for high performance in object detection, and minimizing box loss helps achieve precise localization of objects.

Objectness loss evaluates how well the model distinguishes between the presence and absence of objects in the grid cells. It ensures that the model correctly identifies cells that contain objects. Objectness loss is typically a binary cross-entropy loss applied to the objectness score (confidence score) predicted by the model. The formula is:

$$\text{obj}_{loss} = -\sum_{i=1}^{N} \left(y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\right) \qquad \text{Equation 4.2}$$

where y_i is 1 if an object is present in the cell, and 0 otherwise, and $\hat{p}_i$ is the predicted probability of an object being present. Objectness loss plays a critical role in training the model to differentiate between background and foreground, improving the detection accuracy.

Class loss measures the accuracy of the predicted class probabilities for the detected objects. It ensures that the model correctly classifies each detected object.

Class loss is usually computed using categorical cross-entropy loss applied to the class probabilities. The formula is:

$$\text{cls}_{loss} = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(\hat{p}_{i,c}) \qquad \text{Equation 4.3}$$

where $y_{i,c}$is the true class label, and $\hat{p}_{i,c}$ is the predicted probability for class c.

In this formula, N represents the total number of samples or bounding boxes in the dataset. Each bounding box corresponds to an object or region of interest in an image. C denotes the total number of classes in the classification task, such as different object categories (e.g., car, person, dog, etc.) that the model is trained to detect. Class c refers to a specific category or label within the set of C classes. For example, if the task involves detecting three classes—car, person, and dog—then C = 3, and c can take values corresponding to these specific classes. $y_{i,c}$ is the ground truth label for sample i and class c (1 if the object belongs to class c, otherwise 0), while $\hat{p}_{i,c}$ is the predicted probability for sample i being classified as class c. This loss function penalizes the model when its predicted probabilities differ from the true class labels, thereby driving the model toward better classification accuracy.

Minimizing class loss is essential for achieving high classification accuracy, ensuring that objects are correctly identified.

mAP is a popular evaluation metric for object detection tasks, which combines precision and recall into a single metric. It calculates the average precision (AP) for each class and then computes the mean over all classes.

Precision and Recall:

$$\text{Precision: Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)+False Positives (FP)}} \qquad \text{Equation 4.4}$$

$$\text{Recall: Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)+False Negatives (FN)}} \qquad \text{Equation 4.5}$$

Average Precision (AP): AP is the area under the precision-recall curve for a particular class.

It can be calculated using the following integral over recall:

$$\text{AP} = \int_0^1 \text{Precision(Recall)}d\text{(Recall)} \qquad \text{Equation 4.6}$$

In this equation, d() refers to the infinitesimal change in recall. The integral sums up the product of precision and the change in recall across all recall values, essentially measuring the precision at every possible recall level and aggregating it over the entire range [0, 1]. This calculation provides a comprehensive view of the model's performance across different decision thresholds. In practice, this integral is often approximated using numerical methods or stepwise summation over discrete recall values derived from the model's predictions. In practice, AP is often approximated using a finite number of points on the precision-recall curve:

$$\text{AP} = \sum_{n} (R_n - R_{n-1})P_n \qquad \text{Equation 4.7}$$

where P_n and R_n are the precision and recall at the n-th threshold.

Mean Average Precision (mAP):

$$\text{mAP} = \frac{1}{C}\sum_{c=1}^{C} \text{AP}_c \qquad \text{Equation 4.8}$$

Here, C is the number of classes, and AP is the average precision for class c.

The YOLOv8l model appended with the ECA attention mechanism achieved remarkable results, outperforming other configurations by presenting the lowest training and validation loss at the 933rd epoch. This is visually represented in Figure 4.1, which provides a graphical insight into the loss metrics over the training period. The superior performance of YOLOv8l-ECA is indicative of the efficiency of the ECA mechanism in enhancing the model's focus on relevant features within the examination setting, thereby improving detection accuracy. In contrast, other attention mechanisms, such as SE and CBAM, did not perform as well. One possible reason is the computational complexity introduced by these mechanisms. For example, SE and CBAM involve additional operations like multi-dimensional pooling or complex feature transformations, which may lead to overfitting or difficulty in capturing subtle cheating behaviors in the relatively constrained examination dataset. Furthermore, these mechanisms might overemphasize certain features while neglecting others, reducing the model's ability to generalize across different cheating actions. On the other hand, the ECA mechanism uses a simpler and more efficient design, relying on a one-dimensional convolution that selectively enhances cross-channel interactions without introducing unnecessary computational overhead. This allows ECA to maintain a balance between computational efficiency and feature selectivity, making it more effective for the specific context of cheating detection in examination settings.

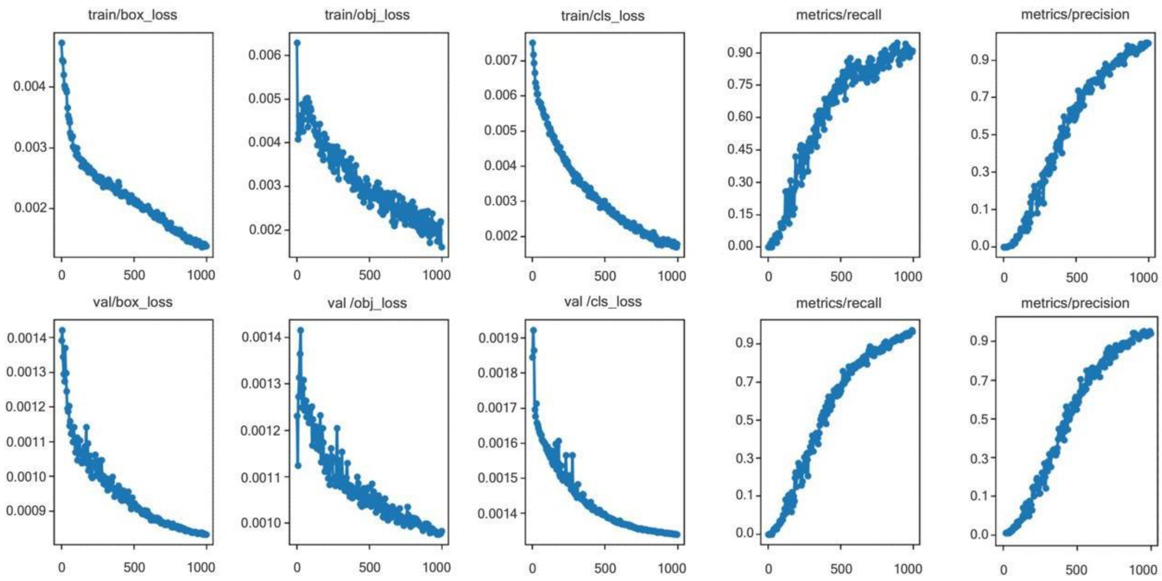

Figure 4.1: Training results of YOLOv8l-ECA

The ECA mechanism's advantage is in its capacity to dynamically adjust the channel-wise feature responses, emphasizing informative features that contribute more significantly to the task of object detection (Dong et al., 2024). This is particularly pertinent in the examination surveillance context, where subtleties in behavior or small objects, such as mobile phones, must be detected with precision from a multitude of distractive background activities.

Interestingly, certain attention mechanism models, when applied to the YOLOv8 model, do not exhibit the same level of effectiveness as the model without attention mechanisms. For example, a comparison between the YOLOv8m model with and without the SE (Squeeze-and-Excitation) attention mechanism revealed that the YOLOv8m model without SE consistently performed better. This was demonstrated by lower error values across all loss categories, including localization loss, objectness loss, and classification loss.

The observed performance difference may be attributed to the SE mechanism's design, which, while effective in other contexts, may not align well with the specific characteristics of the dataset used in this study. This misalignment could result in the overemphasis of certain features or an inability to capture the contextual dependencies required for accurate cheating detection, ultimately hindering the model's performance.

The confusion matrix presented in Figure 4.2 provides further insight into the best model's performance, detailing the classification accuracy across various subcategories of cheating behavior. Notably, the accuracy for detecting the subclass "taking out mobile phones" stands at 84.13%, which highlights the model's ability to recognize this specific form of cheating effectively. This high accuracy may be attributed to the distinctiveness of the "taking out mobile phones" behavior, which likely generates clear and distinguishable patterns in the input data, making it easier for the model to detect. Understanding the accuracy across different subclasses is crucial for evaluating the model's overall utility, as balanced performance across subclasses ensures that no particular type of cheating behavior is overlooked.

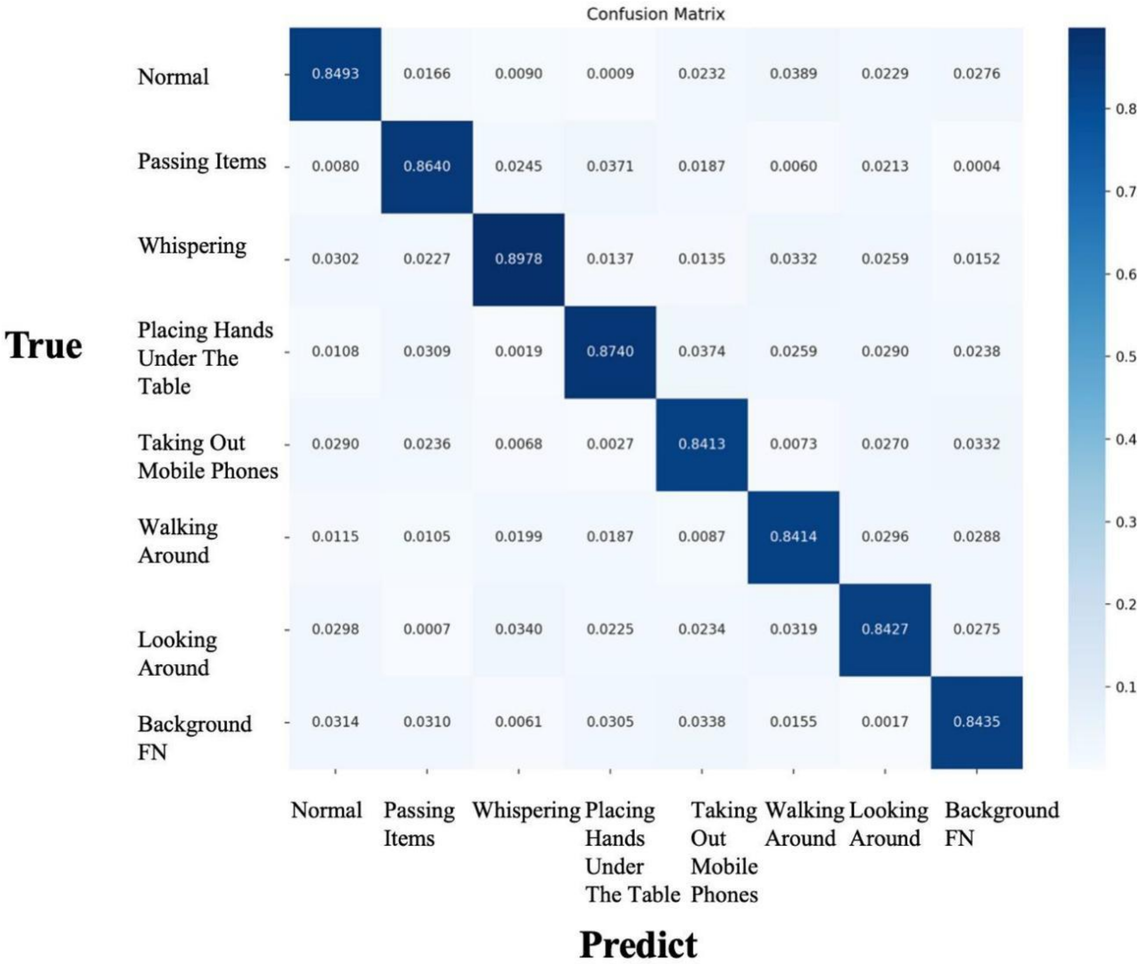

Figure 4.2: Confusion matrix of YOLOv8l-ECA

Normal: The accuracy for predicting 'Normal' behavior is 0.8493, which means the model correctly identified normal behavior 84.93%. Passing Items: The accuracy for 'Passing Items' is 0.8640, indicating that the model correctly predicted this behavior 86.40%. Whispering:The model's accuracy for 'Whispering' is 0.8978, meaning it was correct 89.78% when predicting this behavior. Placing Hands Under The Table: The accuracy here is 0.8740, so the model correctly identified this behavior 87.40%. Taking Out Mobile Phones: The model had an accuracy of 0.8413 for 'Taking Out Mobile Phones', identifying this correctly 84.13%. Walking Around: For 'Walking Around', the accuracy is 0.8414, with the model correctly predicting this behavior 84.14%. Looking Around: Lastly, the 'Looking Around' behavior has an accuracy of 0.8427, which means the model was correct 84.27%. The overall accuracy for each label is quite high, suggesting that the model is performing well across all these behaviors. The high accuracy rate confirms the model's capability to discern between nuanced cheating actions and legitimate exam-taking behavior with a high degree of precision. This effectiveness is attributed to the synergy between the advanced object detection capabilities of the YOLOv8 architecture and the focus-enhancing properties of the ECA mechanism, which together create a powerful tool for monitoring and ensuring examination integrity. Table 4.4 provides a detailed breakdown of the model's performance across various classes of cheating behavior.

Table 4.4: The accuracy of various behaviors.

Behavior	**Accuracy**
Norma	0.8493
Passing Items	0.8640
Whispering	0.8978
Placing Hands Under The Table	0.8740
Taking Out Mobile Phones	0.8413
Walking Around	0.8414
Looking Around	0.8427

Moreover, the mean Average Precision (mAP) of the improved YOLOv8 model stands at a significant 85.67%. The mAP metric is a standard evaluation criterion in object detection tasks that aggregates the precision across different object classes, offering a holistic measure of the model's overall detection capability. A mAP score greater than 80% is indicative of the model's effectiveness in identifying various forms of cheating with a high degree of reliability.

From an analytical perspective, the performance metrics described provide a twofold benefit. Firstly, they illustrate the successful application of an advanced deep learning model to a real-world problem, where the stakes are high, and the cost of errors is significant. Secondly, they provide a foundation for future research, suggesting that attention mechanisms, particularly ECA, have a substantial impact on the model's performance and may be the direction for further enhancement.

The detailed performance analysis also underscores the importance of fine-tuning and selecting the appropriate attention mechanisms for specific tasks. It further opens up discussion on the transferability of these findings, where attention mechanisms validated in one domain of object detection can provide insights into their potential applications in other domains.

In conclusion, the results presented in Table 4.3 and Figures 4.1 and 4.2 affirm the hypothesis that integrating attention mechanisms with YOLOv8 can significantly enhance cheating detection in examination environments. The comprehensive evaluation and comparison of different YOLOv8 configurations with and without attention mechanisms demonstrate the utility of targeted feature refinement in improving object detection performance.

4.5 Testing Analysis

Upon the deployment of the trained YOLOv8l-ECA model on the second video dataset, which was set aside as the test set, we focused primarily on the accuracy of detection. Accuracy in the context of cheating detection is of paramount importance due to the serious implications of both false positives and false negatives. The former can unfairly accuse students of cheating, while the latter can undermine the integrity of the examination process by failing to identify dishonest conduct. Consequently, while speed is an essential metric, ensuring high detection accuracy was the priority, with the goal of achieving an acceptable balance between the two.

The performance results revealed that the improved YOLOv8 model operated at a detection speed of 27 frames per second (FPS), exceeding the minimum threshold of 24 FPS needed to ensure a smooth video browsing experience. This speed is adequate for real- time detection, which is vital during live proctored exams where immediate action may be required upon detecting cheating. Figure 4.3 effectively illustrates this capability, demonstrating that the majority of the people engaged in the simulated exam were detected smoothly without noticeable lag, ensuring that the system's responsiveness meets the practical requirements for real-time applications.

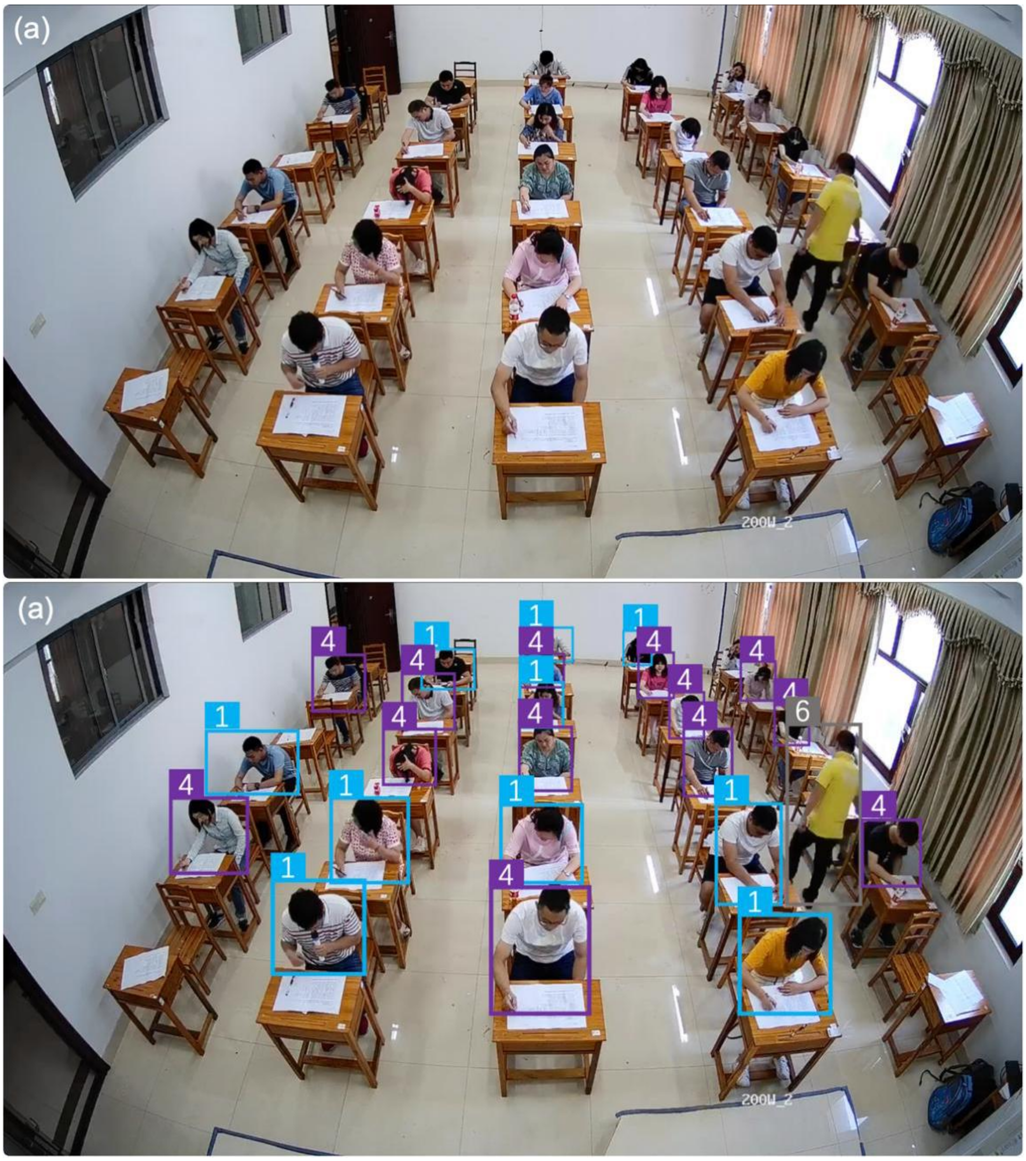
(a)
ZOOM_2
(a)
ZOOM_2

Figure 4.3: Test results of YOLOv8l-ECA for two-event detection (a and 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]

Furthermore, the detection accuracy of the YOLOv8l-ECA model was measured at 82.71% for the labelled test set. This metric is illustrated in the confusion matrix presented in Figure 4.4, which provides a detailed breakdown of the model's performance across various classes of cheating behavior.

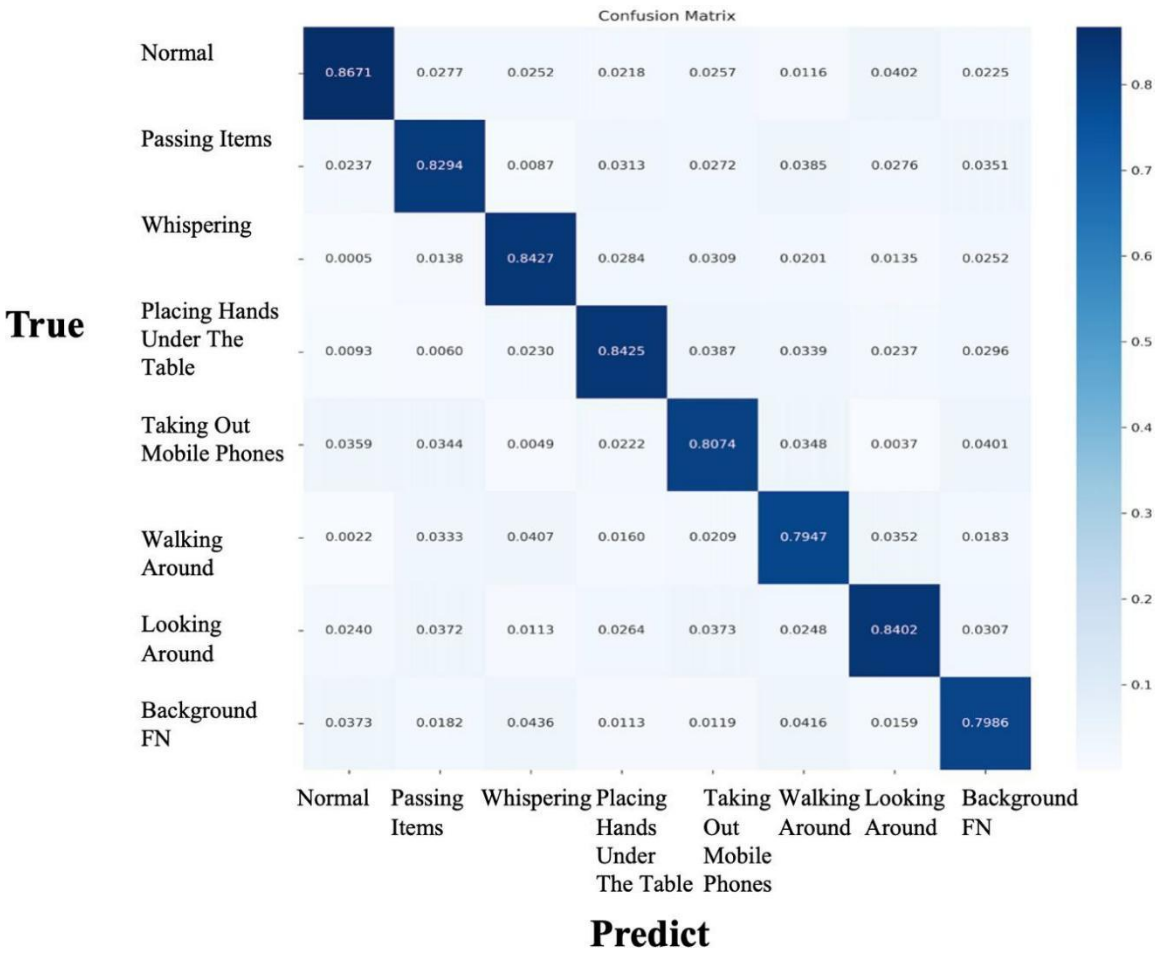

Figure 4.4: Test results of YOLOv8l-ECA

In the comparative analysis of the object detection algorithms, the effectiveness of the YOLOv8l model, particularly when integrated with the ECA mechanism, was not solely due to the inherent strengths of the YOLO architecture but also significantly because of the role played by the attention mechanism. Attention mechanisms in neural networks, by design, allow the model to focus 'attention' selectively on certain parts of the input data which is more informative for the task at hand. This is analogous to how human attention allows us to focus on particular aspects of visual field, such as reading a street sign while driving, despite a barrage of other visual information. In the context of cheating detection, the attention mechanism enables the YOLOv8l model to prioritize visual cues that are more likely indicative of cheating behavior—such as subtle hand movements, glances, or the presence of unauthorized materials—over less significant information in the background.

The enhanced performance of YOLOv8 with attention mechanisms also stems from the ability of these models to reduce the background noise and focus on relevant attributes that contribute to a particular class prediction. For example, the ECA mechanism can effectively give precedence to the texture, shape, and size of small objects within complex visual scenes, thus allowing for more precise and contextually aware detections (Cheng et al., 2023). This is crucial in an examination setting where cheating instances may involve small, easily overlooked details within a broad and dynamic visual frame. The application of attention mechanisms in this case not only improves the accuracy of detection but also ensures that the model remains agile and capable of operating at the desired speed. This synergy of speed and focused detection accuracy is what sets the YOLOv8l-ECA model apart from other detection systems and reaffirms the significance of attention mechanisms in improving the performance of convolutional neural networks for specific and challenging tasks such as cheating detection in examinations.

In experiments, both SSD and Faster R-CNN were configured similarly to YOLOv8 and tested on the same data set for a direct performance comparison. Despite the comparable configuration, the YOLOv8 model integrated with the ECA mechanism exhibited better accuracy, affirming the hypothesis that targeted improvements to a model's attention capabilities can significantly enhance its performance in specific tasks.

When juxtaposed with classical object detection algorithms like SSD and Faster R-CNN, the improved YOLOv8 model's superiority becomes evident, as shown in Table 4.5. While both SSD and Faster R-CNN are well-established algorithms with proven capabilities in various computer vision tasks, they fall short in the context of cheating detection. This can be attributed to their architectural limitations in handling the high variability and subtleties of cheating behaviors as opposed to the straightforward object detection tasks for which they were originally designed.

SSD, known for its balance of speed and accuracy, is generally considered a strong performer in object detection tasks. However, its limitation comes to the fore in specialized tasks such as cheating detection, where the requirement for detecting small or inconspicuous objects under varied lighting conditions and from different angles can impact its performance. Faster R-CNN, although highly accurate in object detection, is constrained by its slower detection speeds, making it less suitable for the real-time aspect of the application at hand.

Table 4.5: Comparison for different target-detection methods.

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

The results analysis also raises important considerations regarding the computational efficiency of the system. While the detection speed is contingent on the computer's configuration, the YOLOv8l-ECA model's ability to maintain a high accuracy on standard computer setups suggests an optimized balance between computational demand and performance. This optimization is crucial for the scalability of the system's deployment across educational institutions with varying levels of computational resources. During testing, a significant majority of labels were accurately identified, with only a minimal number of instances, where labels were either missing or misidentified. In specific applications, it is feasible to determine whether a student is engaging in cheating behavior based on the given context and circumstances.

4.6 Summary

In conclusion, the exploration of results presented in this chapter provides a detailed account of the performance capabilities of an improved YOLOv8 algorithm integrated with an attention mechanism for the purpose of cheating detection in examination environments. The chapter began with an exposition of the dataset, which was painstakingly labeled to train and test the model. With over a thousand annotations covering a spectrum of cheating behaviors, the dataset was foundational in developing a system capable of discerning between cheating and non-cheating actions with high precision.

The experiment setup was articulated, specifying the conditions under which the YOLOv8 algorithm was evaluated. This included a description of the hardware used, the parameters set for the training and validation phases, and the metrics chosen to measure performance. The results section provided a granular view of the algorithm's accuracy, showcasing the model's ability to identify specific cheating actions with a high degree of confidence and within the real-time constraints necessary for practical application.

The subsequent analysis of these results delved into the interpretation of the data, considering the broader implications for automated surveillance systems in academic settings. The improved YOLOv8 model demonstrated a substantial advancement over traditional methods, with the attention mechanism proving instrumental in enhancing the system's focus and accuracy. However, the results also prompted a discussion on the limitations of the current system and outlined potential avenues for future research.

CHAPTER 5

CONCLUSION

5.1 Summary of Research Contributions

The primary contribution lies in the adaptation and customization of the YOLOv8 algorithm for the specific purpose of detecting cheating behaviors in examination settings. Focusing on a domain-specific application of a general object detection framework has advanced the state-of-the-art in automated surveillance for educational settings. The integration of an attention mechanism within YOLOv8 represents a novel approach to enhancing model sensitivity to subtle indicators of dishonest behaviors, which are often overlooked by conventional detection systems.

The creation of a unique, specialized dataset for the detection of cheating in examination settings represents a pivotal contribution of this study. Developing a dataset tailored to the nuanced requirements of this domain involved simulating examination conditions and capturing a wide range of cheating behaviors on video. The meticulous process of generating and labeling this dataset addressed a significant gap in existing research, which often lacks the specificity required for high-fidelity detection in educational settings. By designing and utilizing a dataset that closely mimics real-world examination scenarios, this research has provided a valuable resource that can be used to benchmark and improve upon current cheating detection methods.

Throughout the research, the exploration into the utilization of various attention mechanisms—namely SE, CBAM, ECA, CA, and SOCA—has provided a comprehensive understanding of how such enhancements can bolster the capabilities of the YOLOv8 framework. The study meticulously detailed the process of creating a highly specialized dataset from simulated examination environments, which included a substantial number of images labeled with precise categories of cheating behaviors. This dataset played a critical role in training the models and presented a unique contribution to the field by offering a benchmark for future research in similar domains.

By achieving a high detection accuracy rate of 82.71%, the system significantly reduces the reliance on manual video review and the associated labor costs. This contribution is particularly noteworthy in the context of large-scale examination venues where traditional invigilation methods are both resource-intensive and less effective. The research offers a scalable, accurate, and efficient solution to a prevalent challenge in academic assessment, thus directly contributing to the enhancement of fairness and integrity in examinations globally.

This research made significant strides in achieving its outlined objectives through a series of systematic and innovative approaches. Firstly, the adaptation and customization of the YOLOv8 algorithm were paramount to this study. The model was tailored to specifically identify various cheating behaviors within examination settings, which included the detection of unauthorized materials, unusual movements, and covert communications among students. The customization process involved rigorous training and fine-tuning to ensure the model's responsiveness and effectiveness in recognizing these specific types of misconduct. The result was a sensitive detection system capable of identifying subtle and overt cheating behaviors with a promising accuracy. Secondly, a major contribution of this research was the incorporation of an attention mechanism into the YOLOv8 framework. This integration aimed to enhance the model's ability to detect subtle cues and patterns that are indicative of cheating. Through extensive testing and evaluation, it was demonstrated that the attention mechanism significantly improved the model's performance. The enhancements were particularly notable in terms of increased accuracy and efficiency, as the model was able to focus on relevant parts of the input data, thereby reducing false positives and negatives. The optimization process involved refining the algorithm to ensure it could process and analyze video feeds swiftly, enabling immediate detection and response to cheating incidents. The real-time capability was validated through numerous trials, confirming that the model operates effectively without the need for high-end computational resources.

The research further contributes to the body of knowledge by establishing a balanced approach between detection speed and accuracy. While maintaining an impressive detection speed of 27 FPS, the YOLOv8l model with ECA attention outperformed traditional object detection models, demonstrating that high accuracy need not come at the expense of speed. This balance is critical in examination environments, where the instantaneous detection of cheating is as important as the accuracy of such detection. The ability to detect at a speed that supports real-time analysis without compromising on performance establishes a new standard for automated cheating detection systems.

Moreover, the findings from this study extend beyond the academic sphere, suggesting practical applications in various other settings. In educational environments, this approach promises to strengthen the reliability and fairness of assessments by providing a robust solution to prevent and detect cheating. This is particularly relevant in the current landscape, where the transition to digital learning platforms is burgeoning, and the need for secure remote proctoring tools is ever-increasing.

In corporate settings, the application of the YOLOv8l-ECA model could ensure that training programs are conducted with integrity, preserving the value and credibility of the qualifications awarded. It can also serve as a tool for monitoring compliance with corporate policies and procedures during virtual activities.

Furthermore, the implications for the security industry are significant. The ability to detect suspicious behaviors efficiently through video surveillance can be instrumental in preventing crime and ensuring public safety. By adapting the model to recognize behaviors that are indicative of criminal activity, security agencies could leverage this technology to enhance their surveillance capabilities.

The research has also showcased the versatility and scalability of the YOLOv8l model with ECA attention mechanism. Given the model's performance on standard computing configurations, there is a strong case to be made for its deployment in environments with limited computational resources. This adaptability makes the model an attractive proposition for a wide range of applications and users, potentially driving wider adoption and influencing the development of future detection systems.

In conclusion, this thesis has made substantial contributions to the field of automated cheating detection by delivering an empirically validated, high-performing model that sets a new benchmark in terms of accuracy and speed. It has also opened up new pathways for the application of such technology across various sectors, providing insights that could shape the development of future surveillance and monitoring systems. The study's findings underscore the critical role of data-driven, AI-powered solutions in addressing the challenges of maintaining integrity and fairness in various aspects of digital engagement.

5.2 System Limitation and Future Works

The research conducted in this thesis has successfully demonstrated the use of an enhanced YOLOv8 model with attention mechanisms to detect cheating behavior in examination settings. The challenge lies in the distinction between anomalous yet innocent behavior and genuine instances of cheating. For instance, a student might simply be distracted or engage in movements that are misinterpreted by the model as suspicious. Therefore, there's a vital need for future research to focus on the development of algorithms that can track multitarget behaviors over time, providing a temporal context to actions captured on video. Such tracking would enable the system to construct a behavioral profile for each individual, considering the frequency and type of any abnormal behavior exhibited. This approach would significantly improve the accuracy of cheating detection by allowing the system to distinguish between consistently suspicious behavior and isolated incidents that are not indicative of dishonest intent.

Looking ahead, the incorporation of multitarget tracking technology presents an exciting avenue for future work. This technology can enable continuous monitoring of all individuals in an examination room, capturing their movements and interactions over time. By integrating behavioral prediction, the system would not only flag potential cheating incidents as they occur but also analyze behavior patterns that may suggest premeditated cheating strategies.

Moreover, the exploration of transformer-based models, such as vision transformers that utilize self-attention mechanisms, could greatly enhance the system’s detection capabilities. It is important to note the distinction between general attention mechanisms and self-attention mechanisms. Attention mechanisms broadly refer to techniques that allow a model to focus on the most relevant parts of the input, often by assigning weights to different features or regions of the input data based on their importance. These mechanisms can involve relationships between different parts of the input or between the input and external information. In contrast, self-attention mechanisms are a specific type of attention mechanism where the model computes attention weights internally within the input itself. This means that each element in the input (e.g., a pixel or token) attends to every other element, enabling the model to capture long-range dependencies and contextual relationships within the same input. The self-attention mechanism used in vision transformers allows for a more holistic understanding of spatial and contextual patterns in images, making it particularly suited for complex detection tasks.

The current system's dependence on visual data alone can be seen as a significant limitation, as cheating behaviors can also manifest in non-visual forms. Incorporating audio data could provide additional context, enabling the detection of whispered conversations, the sound of typing on a keyboard, or the rustling of papers, which are often associated with cheating. These audio cues can be critical in environments where visual monitoring might miss subtle but significant cheating activities. By integrating audio sensors, the system could better identify and interpret these sounds, enhancing the overall accuracy and reliability of cheating detection. Additionally, the use of thermal imaging could reveal the use of electronic devices that may not be visible to standard cameras but emit heat. For instance, students might conceal devices under their clothing or desks, which would be detectable through their heat signatures. This multi-modal approach, combining visual, audio, and thermal data, would create a more comprehensive surveillance system capable of detecting a broader range of cheating behaviors and thereby improving the integrity of the examination process.

Furthermore, the model's performance could be significantly enhanced by improving the robustness of the dataset. While the current dataset is sufficient, it may not cover all possible forms of cheating or examination environments. Future research should aim to expand the dataset to include a wider range of scenarios, including different room layouts, student behaviors, and potential cheating devices. For example, students might develop new ways to cheat that have not yet been documented or included in the dataset, such as using advanced miniature electronic devices or exploiting new technological loopholes. Collecting data across various educational institutions and examination types would improve the model's generalizability and reliability. This broader and more diverse dataset would enable the system to better recognize and adapt to new cheating methods as they emerge, thereby maintaining its effectiveness over time.

Another crucial consideration for future research is the system's scalability and accessibility. While the YOLOv8 model runs efficiently on regular performance computers, there may be significant variations in computing resources across different educational settings. Some institutions might have access to high-end servers and advanced computational tools, while others may rely on more modest computing devices commonly used in everyday educational environments. Future work could address these disparities by optimizing the model to run on a broader range of hardware. This would involve developing lightweight versions of the model that can operate effectively on less powerful machines without compromising on performance. By ensuring that the model is accessible and usable across a wide spectrum of hardware capabilities, the system could be more widely adopted and implemented, thereby enhancing its overall impact on academic integrity.

In summary, while the current research has made significant strides in the field of automated cheating detection, there remains ample opportunity for further enhancement. Future work should aim to create a more sophisticated, context-aware system that leverages multitarget tracking, integrates multiple data sources, and utilizes the latest advancements in AI and machine learning to ensure fair and accurate examination processes. The evolution of this research will contribute to a more secure academic environment and uphold the standards of educational excellence.

REFERENCES

Andrew, G. H., Zhang, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications*. arXiv. https://doi.org/10.48550/arXiv.1704.04861

Aqeel, Z., Jamil, M., & Shahzad, M. (2023). A novel CNN-RNN model for e-cheating detection based on video surveillance. *University of Wah Journal of Computer Science, 5*(1), 1–13. https://uwjcs.org.pk/index.php/ojs/article/view/64

Asadullah, M., & Nisar, S. (2016). An automated technique for cheating detection. In *2016 Sixth International Conference on Innovative Computing Technology (INTECH)* (pp. 251–255). IEEE. https://doi.org/10.1109/INTECH.2016.7845032

Ashish, V., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., N. Gomez, A., & Polosukhin, I. (2017). *Attention is all you need*. arXiv. https://doi.org/10.48550/arXiv.1706.03762

Baron, N. S. (2017). Reading in a digital age. *Phi Delta Kappan, 99*(2), 15–20. https://doi.org/10.1177/0031721717734184

Carranza-García, M., Torres-Mateo, J., Lara-Benítez, P., & García-Gutiérrez, J. (2020). On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing, 13*(1), 89. https://doi.org/10.3390/rs13010089

Christopher, W., & Poore, J. (2022). Developing imperceptible adversarial patches to camouflage military assets from computer vision enabled technologies. *arXiv*. https://doi.org/10.48550/arXiv.2202.08892

Condell, O., Gubbels, S., Nielsen, J., Espenhain, L., Frimodt-Møller, N., Engberg, J., & Kristensen, B. (2016). Automated surveillance system for hospital-acquired urinary tract infections in Denmark. *Journal of Hospital Infection, 93*(3), 290–296. https://doi.org/10.1016/j.jhin.2016.04.001

Dai, J. B., Long, M. L., Zhao, H. W., & Chen, F. J. (2012). Algorithm of the exam abnormal behavior detection. *Computer Engineering and Applications, 42*, 236–240.

Dai, T., Cai, J., Zhang, Y., Xia, S. T., & Zhang, L. (2019, June 15–20). Second-order attention network for single image super-resolution. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11057–11066). IEEE. https://doi.org/10.1109/CVPR.2019.01132

Deepa, R., Tamilselvan, E., Abrar, E. S., & Sampath, S. (2019, April 4–6). Comparison of YOLO, SSD, Faster RCNN for real time tennis ball tracking for action decision networks. In *2019 International Conference on Advances in Computing and Communication Engineering (ICACCE)*. https://doi.org/10.1109/ICACCE.2019.8697256

Dhar, S., & Pal, S. (2022). Surface reconstruction: Roles in the field of computer vision and computer graphics. *International Journal of Image and Graphics, 22*(01), 2250008. https://doi.org/10.1142/S0219467822500084

Dillon, R., Jordan, K., Howie, J., & Dawood, A. (2023). Real-time flying object detection with YOLOv8. *arXiv*. https://doi.org/arxiv-2305.09972

Dong, L., Wei, F., Zhou, M., & Xu, K. (2015). Question answering over Freebase with multi-column convolutional neural networks. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. https://www.semanticscholar.org/paper/Question-Answering-over-Freebase-with-Multi-Column-Dong-Wei/1ef01e7bfab2041bc0c0a56a57906964df9fc985

Du, L., Zhang, R., & Wang, X. (2020). Overview of two-stage object detection algorithms.

Journal of Educational Psychology, 31(1), 54–64.

El Kohli, S., Jannaj, Y., Maanan, M., & Rhinane, H. (2022). Deep learning: New approach for detecting scholar exams fraud. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-4/W3-2021*, 103–107. https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-103-2022

Fang, M.-T., Chen, Z.-J., Przystupa, K., Li, T., Majka, M., & Kochan, O. (2021). Examination of abnormal behavior detection based on improved YOLOv3. *Electronics, 10*(2), 197. https://doi.org/10.3390/electronics10020197

Faucher, D., & Caves, S. (2009). Academic dishonesty: Innovative cheating techniques and the detection and prevention of them. *Teaching and Learning in Nursing, 4*(1), 37–41. https://doi.org/10.1016/j.teln.2008.09.003

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

Gudmundsson, I., & Falco, G. (2022). Porting computer vision models to the edge for smart city applications: Enabling autonomous vision-based power line inspection at the smart grid edge for unmanned aerial vehicles (UAVs). *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS)*. https://doi.org/10.24251/HICSS.2022.929

Guo, K., Zou, D., & Chen, X. (2016). 3D mesh labeling via deep convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).*

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017, October 22–29). Mask R-CNN. In

2017 IEEE International Conference on Computer Vision (ICCV) (pp. 2980–2988). IEEE. https://doi.org/10.1109/ICCV.2017.322

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

Hossain, Z. T., Roy, P., Nasir, R., Nawsheen, S., & Hossain, M. I. (2021, December 3–4). Automated online exam proctoring system using computer vision and hybrid ML classifier. In *2021 IEEE International Conference on Robotics, Automation, Artificial Intelligence and Internet-of-Things (RAAICON)* (pp. 14–17). IEEE. https://doi.org/10.1109/RAAICON51346.2021.9420629

Hou, Q., Zhou, D., & Feng, J. (2021, June 20–25). Coordinate attention for efficient mobile network design. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 13708–13717). IEEE. https://doi.org/10.1109/CVPR46437.2021.01350

Hu, H., Guan, Q., Chen, S., Ji, Z., & Lin, Y. (2017). Detection and recognition for life state of cell cancer using two-stage cascade CNNs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17*(3), 887–898. https://doi.org/10.1109/TCBB.2017.2670545

Hu, J., Shen, L., & Sun, G. (2018, June 18–23). Squeeze-and-excitation networks. In *2018*

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7132–7141). IEEE. https://doi.org/10.1109/CVPR.2018.00745

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2020.2995783

Hussein, F., Al-Ahmad, A., El-Salhi, S., Alshdaifat, E. A., & Al-Hami, M. T. (2022). Advances in contextual action recognition: Automatic cheating detection using machine learning techniques. *Data, 7*(9), 122. https://www.mdpi.com/2306-5729/7/9/122

Jackson, D. F. (1961). The detection of cheating in multiple choice tests. *British Journal of Educational Psychology, 31*(2), 132–138. https://doi.org/10.1111/j.2044-8279.1961.tb02408.x

Jiang, A., Nacenta, M. A., & Ye, J. (2022). Visualizations as intermediate representations (VLAIR): An approach for applying deep learning-based computer vision to non-image-based data. *Visual Informatics, 6*(3), 35–50. https://doi.org/10.1016/j.visinf.2022.05.001

Jocher, G., Qiu, J., & Chaurasia, A. (2023). *Ultralytics YOLO (Version 8.0.0)* [Computer software]. GitHub. https://github.com/ultralytics/ultralytics

Johnson, R., & Zhang, T. (2015). Semi-supervised convolutional neural networks for text categorization via region embedding. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS)*, Montreal, Canada.

Joseph, R., & Ali, F. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. https://doi.org/10.48550/arXiv.1804.02767

Joseph, R., Dary, S., Redmon, R., & Ali, F. (2015). You only look once: Unified, real-time object detection. *arXiv preprint arXiv:1506.02640*. https://doi.org/10.48550/arXiv.1506.02640

Kamalov, F., Sulieman, H., & Santandreu Calonge, D. (2021). Machine learning based approach to exam cheating detection. *PLOS ONE, 16*(8), e0254340. https://doi.org/10.1371/journal.pone.0254340

Kaye, A. D., Okeagu, C. N., Pham, A. D., Silva, R. A., Hurley, J. J., Arron, B. L., & Cornett, E. M. (2021). Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives. *Best Practice & Research Clinical Anaesthesiology*, *35*(3), 293–306. https://doi.org/10.1016/j.bpa.2020.11.009

Khan, S., Rahmani, H., & Shah, S. A. A. (2018). *A guide to convolutional neural networks for computer vision*. Morgan & Claypool Publishers.

Kingma, D. P. (2014). *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Kontogianni, T. (2022). *Object discovery, interactive and 3D segmentation for large-scale computer vision tasks* [Doctoral dissertation, RWTH Aachen University]. RWTH Publications. https://publications.rwth-aachen.de/record/844228

Kornaros, G. (2022). Hardware-assisted machine learning in resource-constrained IoT environments for security: Review and future prospective. *IEEE Access, 10,* 58603–58622. https://doi.org/10.1109/ACCESS.2022.3178857

Krishnamurthy, S. (2020). The future of business education: A commentary in the shadow of the COVID-19 pandemic. *Journal of Business Research, 117*, 1–5. https://doi.org/10.1016/j.jbusres.2020.05.034

Lamping, C., Derks, M., Groot Koerkamp, P., & Kootstra, G. (2022). ChickenNet – An end-to-end approach for plumage condition assessment of laying hens in commercial farms using computer vision. *Computers and Electronics in Agriculture, 194*, 106695. https://doi.org/10.1016/j.compag.2022.106695

Laurisa, F. O., Sajul, K. M., & Tirtana, A. (2022). Cheating detection during exam with YOLO V4. *IC-ITECHS, 3*(1), 231–235.

Li, P.-W., Dai, J.-Y., & Ying, J. (2020). Vehicle detection method based on improved Faster R-CNN. *Journal of Nanchang Hangkong University (Natural Science Edition), 34*(3), 73–79. https://doi.org/10.3969/j.issn.2096-8566.2020.03.011

Li, W., Liu, K., Zhang, L., & Cheng, F. (2020). Object detection based on an adaptive attention mechanism. *Proceedings of International Conference on Artificial Intelligence and Computer Engineering*. https://www.nature.com/articles/s41598-020-67529-x

Li, Y., Padmanabhan, A., Zhao, P., Wang, Y., Xu, G. H., & Netravali, R. (2020, July). Reducto: On-camera filtering for resource-efficient real-time video analytics. In *Proceedings of the ACM SIGCOMM 2020 Conference* (pp. 359–376). https://doi.org/10.1145/3387514.3405863

Li, Z., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., Xu, D., & 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., & Dollár, P. (2017, October 22–29). Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2999–3007). IEEE. https://doi.org/10.1109/ICCV.2017.324

Lin, Y., Wu, X., & Li, H. (2015). Cheating behavior detection in examination room based on background subtraction. *Journal of University of Jinan, 29*, 406–409.

Lipson, S., & Karthikeyan, L. (2016). The art of cheating in the 21st millennium: Innovative mechanisms and insidious ploys in academic deceit. *International Journal of Education, 8*(2), 48–72.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *Computer Vision – ECCV 2016* (pp. 21–37). Springer. https://doi.org/10.1007/978-3-319-46448-0_2

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Lu, S., Wang, B., Wang, H., Chen, L., Min, L., & Zhang, X. (2019). A real-time object detection algorithm for video. *Computers & Electrical Engineering, 77*, 398–408. https://doi.org/10.1016/j.compeleceng.2019.05.009

Ma, B., Fu, Z., Rakheja, S., Zhao, D., He, W., Ming, W., & Zhang, Z. (2024). Distracted driving behavior and driver's emotion detection based on improved YOLOv8 with attention mechanism. *IEEE Access, 12*, 37983–37994. https://doi.org/10.1109/ACCESS.2024.3374726

Marco Tulio, R., Sameer, S., & Carlos, G. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *arXiv preprint arXiv:1602.04938*. https://doi.org/10.48550/arXiv.1602.04938

Mariano, V. Y., Junghye, M., Jin-Hyeong, P., Kasturi, R., Mihalcik, D., Huiping, L., & Drayer, T. (2002, August 11–15). Performance evaluation of object detection algorithms. In *2002 International Conference on Pattern Recognition* (Vol. 3, pp. 965–969). IEEE. https://doi.org/10.1109/ICPR.2002.1048191

Max, J., Andrea, V., & Andrew, Z. (2014). Speeding up convolutional neural networks with low rank expansions. *arXiv preprint arXiv:1405.3866*.

https://doi.org/10.48550/arXiv.1405.3866

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

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), e3292. https://doi.org/10.1002/rev3.3292

Nascimento, J. C., & Marques, J. S. (2006). Performance evaluation of object detection algorithms for video surveillance. *IEEE Transactions on Multimedia, 8*(4), 761–774. https://doi.org/10.1109/TMM.2006.878958

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

Odongo, D. A., Agyemang, E., & Forkuor, J. B. (2021). Innovative approaches to cheating: An exploration of examination cheating techniques among tertiary students. *Education Research International, 2021*, 1–7. https://doi.org/10.1155/2021/6663832

Padhiyar, P., Parmar, K., Parmar, N., & Degadwala, S. (2023, April 26–28). Visual distance fraudulent detection in exam hall using YOLO detector. In *2023 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICICT57944.2023.10121725

Padilla, R., Netto, S. L., & Silva, E. A. B. D. (2020, July 1–3). A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 237–242). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145182

Potluri, T., & Sistla, V. P. K. (2023). MG-Net: Multiple person and gadget detection for online exam proctoring system. In *Proceedings of the Third International Conference on Advances in Computer Engineering and Communication Systems* (pp. 445–456). Springer, Singapore. https://doi.org/10.1007/978-981-99-2671-4_35

Raghunandan, A., Mohana, Raghav, P., & Aradhya, H. V. R. (2018, April 3–5). Object detection algorithms for video surveillance applications. In *2018 International Conference on Communication and Signal Processing (ICCSP)* (pp. 563–568). IEEE. https://doi.org/10.1109/ICCSP.2018.8524254

Ren, J., & Wang, Y. (2022). Overview of object detection algorithms using convolutional neural networks. *Computational Intelligence and Neuroscience*, *2022*, 1–13. https://doi.org/10.1155/2022/8754343

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

Resta, P., & Laferrière, T. (2008). Issues and challenges related to digital equity. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 765–778). Springer US. https://doi.org/10.1007/978-0-387-73315-9_44

Sainath, T. N., Kingsbury, B., Saon, G., Soltau, H., Mohamed, A.-R., Dahl, G., & Ramabhadran, B. (2015). Deep convolutional neural networks for large-scale speech tasks. *Neural Networks, 64*, 39–48. https://doi.org/10.1016/j.neunet.2014.08.005

Schockaert, S., De Cock, M., & Kerre, E. E. (2010). Reasoning about fuzzy temporal information from the web: Towards retrieval of historical events. *Soft Computing, 14*(8), 869–886. https://doi.org/10.1007/s00500-009-0471-8

Sharma, N. K., Gautam, D. K., Rathore, S., & Khan, M. R. (2021). WITHDRAWN: CNN implementation for detect cheating in online exams during COVID-19 pandemic: A CVRU perspective. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.05.490

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

Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2015). Multi-scale convolutional neural networks for lung nodule classification. In *Information Processing in Medical Imaging* (pp. 588–599). Springer.

Solawetz, J. (2020, July 13). What are anchor boxes in object detection? [Blog post]. *Roboflow*. https://blog.roboflow.com/what-is-an-anchor-box/

Su, H., Maji, S., Kalogerakis, E., & Learned-Miller, E. (2015, December 7–13). Multi-view convolutional neural networks for 3D shape recognition. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 945–953). IEEE. https://doi.org/10.1109/ICCV.2015.114

Sun, W., Dai, L., Zhang, X., Chang, P., & He, X. (2022). RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring. *Applied Intelligence, 52*, 8448–8463. https://doi.org/10.1007/s10489-021-02893-3

Tabian, I., Fu, H., & Sharif Khodaei, Z. (2019). A convolutional neural network for impact detection and characterization of complex composite structures. *Sensors, 19*(22), 4933. https://doi.org/10.3390/s19224933

Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Medical Informatics and Decision*

Making, 21, 324. https://doi.org/10.1186/s12911-021-01691-8

Ting, Q., Hansong, S., Gaohua, L., & Meng, W. (2019). Object detection algorithm based on improved feature extraction network. *Laser & Optoelectronics Progress*. Trabelsi, Z., Parambil, M. M. A., Alnajjar, F., & Ali, L. (2023). Behavioral-based real-time cheating detection in academic exams using deep learning techniques. *AIP Conference Proceedings, 2909*(1), 020001. https://doi.org/10.1063/5.0181921

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., & Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access, 7,* 65579–65615. https://doi.org/10.1109/ACCESS.2019.2916648

Vercauteren, T., Unberath, M., Padoy, N., & Navab, N. (2020). CAI4CAI: The rise of contextual artificial intelligence in computer-assisted interventions. *Proceedings of the IEEE, 108*(1), 198–214. https://doi.org/10.1109/JPROC.2019.2948771

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, June 13–19). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11531–11539). IEEE. https://doi.org/10.1109/CVPR42600.2020.01155

Wang, X., Gao, H., Jia, Z., & Li, Z. (2023). BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors, 23*(20), 8361. https://www.mdpi.com/1424-

8220/23/20/8361

Wang, Z., Chen, J., Yu, P., Feng, B., & Feng, D. (2024). SC-YOLOv8 network with soft-pooling and attention for elevator passenger detection. *Applied Sciences*, *14*(8), 3321. https://www.mdpi.com/2076-3417/14/8/3321

Wei, Z., Chang, M., & Zhong, Y. (2023). Fruit freshness detection based on YOLOv8 and SE attention mechanism. *Academic Journal of Science and Technology*, 6(1), 195-197.

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Computer Vision – ECCV 2018* (pp. 3–19). Springer, Cham. 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: A review. *Neurocomputing, 396,* 39–64. https://doi.org/10.1016/j.neucom.2020.01.085

Xu, T., & Xiao, H. (2021). Application of SSD core detection algorithm in intelligent visual monitoring of examination room. *Journal of Physics: Conference Series*, *2037*(1), 012085. https://doi.org/10.1088/1742-6596/2037/1/012085

Yadav, N., & Binay, U. (2017). Comparative study of object detection algorithms. *International Journal of Computer Applications, 162*(1), 34–39. https://doi.org/10.5120/ijca2017913274

Yang, G., Wang, J., Nie, Z., Yang, H., & Yu, S. (2023). A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy, 13*(7), 1824. https://www.mdpi.com/2073-4395/13/7/1824

Yongzheng, L., Xiaoyu, W., Hengjian, L., & Zhongguang, L. (n.d.). Cheating behavior detection in examination room based on background subtraction. *Journal of*

University of Jinan (Science and Technology).

Yoon, K. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*. https://doi.org/10.48550/arXiv.1408.5882

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://www.mdpi.com/2072-666X/13/7/1058

Zhang, Q., Wan, C., & Han, W. (2019). A modified faster region-based convolutional neural network approach for improved vehicle detection performance. *Multimedia Tools and Applications, 78*, 29431–29446. https://doi.org/10.1007/s11042-019-7624-z

Zhao, R., Niu, X., Wu, Y., Luk, W., & Liu, Q. (2017). Optimizing CNN-based object detection algorithms on embedded FPGA platforms. In *Applied Reconfigurable Computing* (pp. 255–267). Springer. https://doi.org/10.1007/978-3-319-56258-2_21

APPENDICES

Pseudo-code

Initialize YOLOv8 Network (Darknet or other backbone) Initialize SE, CBAM, ECA, CA, and SOCA attention modules

for epoch in 1...N:

for batch in DataLoader: images, ground_truths = batch

// Forward pass through initial YOLOv8 layers feature_maps = YOLOv8_Backbone(images)

/ Squeeze-and-Excitation (SE) Attention for fm in feature_maps:

SE_out = SE_Block(fm) fm = fm * SE_out

// Convolutional Block Attention Module (CBAM) Attention

for fm in feature_maps: CBAM_out = CBAM_Block(fm) fm = fm * CBAM_out

// Efficient Channel Attention (ECA) Attention for fm in feature_maps:

ECA_out = ECA_Block(fm) fm = fm * ECA_out

// Coordinate Attention (CA) Attention for fm in feature_maps:

CA_out = CA_Block(fm) fm = fm * CA_out

// Second-order Attention Network (SOCA) Attention for fm in feature_maps:

SOCA_out = SOCA_Block(fm) fm = fm * SOCA_out

// Continue with standard YOLOv8 object detection and classification

detections = YOLOv8_Head(feature_maps)

// Calculate Loss

loc_loss = MSE(detections.bboxes, ground_truths.bboxes) obj_loss = MSE(detections.obj_scores, ground_truths.obj_labels)

class_loss = CrossEntropy(detections.class_probs, ground_truths.class_labels) total_loss = lambda_loc * loc_loss + lambda_obj * obj_loss + lambda_class * class_loss

// Backward pass Backpropagate(total_loss)

// Update weights UpdateWeights()

// End of one epoch Validate_on_ValidationSet()

YOLOv8_Backbone(images): This function performs the forward pass of the image through the initial YOLO layers and returns the feature maps at different stages.

SE_Block, CBAM_Block, ECA_Block, CA_Block, SOCA_Block: These functions represent the forward pass through each type of attention block. They take a feature map as input and return an attention map which is then element-wise multiplied with the original feature map.

YOLOv8_Head(feature_maps): This function performs object detection and classification on the feature maps (post-attention).

ComputeLoss: This function computes the loss between the predicted detections and ground truth.

Backpropagate and UpdateWeights: These functions perform the backward pass and update the model weights.