Online Classroom Student Engagement Analysis using Enhanced YOLOv5

Shuai Wang
Institute of Computer Science and Digital Innovation, UCSI University, No.1, Jalan Menara Gading, Cheras, Kuala Lumpur, Malaysia
Abdul Samad Shibghatullah
Institute of Computer Science and Digital Innovation, UCSI University, No.1, Jalan Menara Gading, Cheras, Kuala Lumpur, Malaysia
Javid Iqbal
Institute of Computer Science and Digital Innovation, UCSI University, No.1, Jalan Menara Gading, Cheras, Kuala Lumpur, Malaysia
Kay Hooi Keoy
UCSI Graduate Business School, UCSI University, No.1, Jalan Menara Gading, Cheras, Kuala Lumpur, Malaysia
E-mail: 1002268166@ucsiuniversity.edu.my,abdulsamad@ucsiuniversity.edu.my,javid@ucsiuniversity.edu.my, keoykh@ucsiuniversity.edu.my
www.ucsiuniversity.edu.my

Abstract

The surge in online education has underscored the pressing issue of cyberbullying in virtual classrooms. This paper introduces an inventive method for early cyberbullying detection by analyzing students' engagement and emotional responses in online classrooms. The proposed *SFER-YOLOv5* model, integrating Student Facial Expression Recognition with an enhanced YOLOv5 object detection model, incorporates transformative optimizations. These include Soft NMS for Non-Maximum Suppression, the integration of a Channel Attention (CA) module within the YOLOv5 architecture, and the use of Enhanced Intersection over Union (EIOU) as the bounding box regression loss function. This approach identifies diminished engagement and emotional irregularities, offering a proactive framework for mitigating cyberbullying in online classrooms.

Keywords: Cyberbullying Mitigation, Emotional Analysis, Facial Expression Recognition, Online Classroom.

1. Introduction

With the rapid advancement of information technology and the global rise of online education, online teaching platforms have become an integral part of modern education. The sudden outbreak of the COVID-19 pandemic in 2020 has had a significant impact on our learning, work, and daily lives. Traditional face-to-face teaching and communication between students and teachers have become a major challenge. Many universities have chosen online teaching as an alternative. However, accompanying this advancement are a host of challenges, one of which is the issue of cyberbullying in the digital realm. Cyberbullying, also known as online harassment or cyber aggression, refers to the malicious use of digital technology to target others with acts of aggression, insults, threats, and more. Particularly in online classroom environments, both students and educators may find themselves vulnerable to cyberbullying, which can severely impact their emotional well-being, psychological state, and overall learning experience.

In the online classroom environment, students and teachers communicate through screens, and teachers can only judge students' listening status based on their facial expressions. Analyzing changes in students' facial expressions can help teachers better understand their listening status and make timely adjustments to the teaching mode [1]. In 1971, American scholar Ekman et al. [2] conducted extensive experiments on facial expressions, categorizing them into six primary emotions: happiness, surprise, fear, sadness, disgust, and anger. Students' emotions in class can be categorized as positive, negative, or neutral. Positive emotions include happiness and surprise. When students show positive emotions, it indicates their willingness to accept the knowledge taught in class, and they are actively engaged in listening and thinking. Negative emotions include sadness, anger, fear, and disgust, which indicate that students have a dislike or lack of attentiveness towards the knowledge taught by the teacher. This can be an indication of their lack of concentration in class. Neutral emotions suggest an average level of student engagement in listening during class. Therefore, analyzing students' facial expressions in online classrooms holds great significance.

Faced with this challenge, many educational institutions and online teaching platforms are turning their attention to how to effectively detect and address cyberbullying in virtual classrooms. Facial expression recognition technology has garnered significant attention due to its potential applications in emotion analysis and affective state monitoring. Currently, there is limited research on facial expression recognition based on YOLOv5. This paper proposes a facial expression recognition method based on an improved YOLOv5 model, aiming to achieve timely detection and intervention of cyberbullying incidents in online classrooms. The original YOLOv5 algorithm achieved an overall recognition accuracy of 73.1% and 83.4% on the Fer2013 dataset and a self-constructed dataset for the three expressions involved in the study: happiness, sadness, and neutral. Through improvements in NMS, and the addition of an attention mechanism module, the improved YOLOv5 algorithm further enhances the recognition accuracy on the Fer2013 dataset and the self-constructed dataset.

2. Related work

2.1. *Online classroom cyberbullying*

Online education has become increasingly prevalent in recent years, especially due to the COVID-19 pandemic, which has posed new challenges and opportunities for educators and students. One of the major challenges is cyberbullying [3], [4], which refers to various forms of online harassment, such as verbal abuse, threats, spreading false information, and exclusion from online discussions, that occur within virtual classroom environments. Cyberbullying can have detrimental effects on students' learning outcomes, psychological well-being, and social relationships [5]. Therefore, it is imperative to develop effective strategies and mechanisms to detect and prevent cyberbullying incidents in online classrooms. Previous research on this topic has adopted various techniques and perspectives, such as behavior analysis, natural language processing, social network analysis, machine learning, and emotion recognition, to identify and intervene in potential bullying situations that students may face in virtual classrooms. Moreover, some studies have highlighted the role of educational interventions and awareness-raising among educators and students about cyberbullying, aiming to foster a safer and more positive online learning environment [6].

2.2. *YOLOv5 and object detection*

In May 2020, UltralyticsLLC released YOLOv5 [7], a lightweight model with an image inference speed of up to 0.007 second, developed using python. It can process 140 FPS and can meet the real-time requirement for video sequences. The four network structures Yolov5s\5m\5l\5x are basically same in principle and content, controlled respectively by the parameters of width_multiple and depth_multiple in width and depth. The size and accuracy of the four versions of the model increase sequentially. In practical applications, the appropriate size of the model can be selected based on different scenarios. The YOLOv5s model has small depth and fewer parameters, making it more applicable to real-time tasks in facial expression recognition due to its faster inference speed compared to the other three versions. Since the release of YOLOv5, its versions have been updated and iterated, and this paper is based on the version 6.0 for improvement.

In summary, prior research has highlighted the urgency of addressing online classroom cyberbullying and the potential of facial expression recognition technology for emotion analysis. Additionally, the advancements in object detection methods, particularly YOLOv5, provide a solid foundation for integrating object detection into cyberbullying detection frameworks. This study builds upon these foundations to propose a novel model *SFER-YOLOv5* for detecting cyberbullying incidents in online classrooms using an enhanced YOLOv5 model for facial expression recognition.

3. Methodology

Currently, there is limited research on facial expression recognition based on YOLOv5, and this method presents a unique opportunity for enhancing online classroom engagement and security. The original YOLOv5 algorithm achieved an overall recognition accuracy of 72% and 83.1% for the three facial expressions of happiness, sadness, and neutrality on the Fer2013 dataset and the self-built dataset. Through improvements in the non-maximum suppression (NMS) algorithm, loss function, and the addition of an attention mechanism module, the improved YOLOv5 algorithm achieved further improvement in recognition accuracy on the Fer2013 dataset and the self-built dataset.

3.1. *Data collection and processing*

As students' understanding of knowledge can be broadly categorized into three levels: mastery, confusion, and non-mastery, in order to gain a more accurate understanding of the facial expression feedback corresponding to these three levels, we conducted a survey in the form of a questionnaire to investigate the relationship between mastery levels and facial expression feedback. The survey involved higher vocational schools, and undergraduate institutions, with a total of 520 questionnaires collected (male: 287, female: 233). Seven basic expressions were selected, including anger, surprise, disgust, happiness, fear, sadness, and neutral.

This study used the Fer2013 dataset and a self-constructed dataset of student facial expressions in the classroom. The Fer2013 dataset contains some non-facial images and mislabeled images. To improve the experimental accuracy, the dataset was carefully selected, cleaned, and filtered, resulting in a total of 11,000 images. The self-constructed dataset of student facial expressions in the classroom was collected through video monitoring, and it consists of 668 images. The images were annotated by extracting frames from the video footage.

3.2. *Enhanced YOLOv5*

3.2.1. *Improved Non-Maximum Suppression (NMS)*

Both NMS and Soft NMS utilize predicted classification confidence as a measure, where higher confidence indicates more accurate localization. In this study, the improved Soft NMS is used instead of the original NMS, effectively improving the performance of the detection.

3.2.2. *Coordinate attention module*

In this study, we introduce the Coordinate Attention (CA) mechanism, which incorporates positional information into channel attention. Experimental results reveal that incorporating the CA attention mechanism module into the network backbone improves detection accuracy to some extent, particularly in classroom settings with numerous students or when students in the back rows are at a considerable distance from the camera. This enhancement is attributed to the CA module (Fig. 1), which reinforces channel features in the feature map, enabling the network to acquire more detailed and effective information.

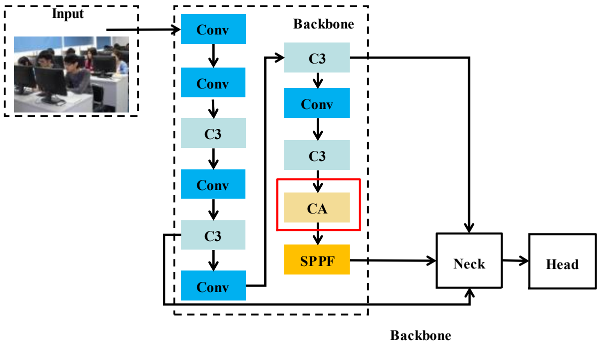

Fig. 1 The improved network structure.

3.2.3. *Improved Loss function*

We have improved the initial loss function by incorporating the Efficient-IOU (EIOU) to compute width and height losses, replacing the aspect ratio calculation based on the CIOU foundation. Furthermore, Focal Loss has been introduced concurrently to mitigate the challenge posed by sample imbalance.

4. Experiments and results

The experimental platform in this study was a 64-bit Windows 10 operating system with a Gen Intel Core i5-11400H CPU and NVIDIA RTX 3050 graphics card. The deep learning framework used was PyTorch, with a development environment consisting of PyTorch 1.8 and Python 3.7. The number of training epochs was set to 300, and the dataset was divided into training, validation, and testing sets following a 6:2:2 ratio. To effectively evaluate the experimental results, a comparative experiment was conducted on the self-constructed dataset and the Fer2013 dataset before and after algorithm improvement. The evaluation criteria were detection accuracy (P) and mean average precision (mAP@0.5).

Table 1 presents the experimental outcomes on our self-constructed dataset. Fig.2 illustrates the detection accuracy (*P*) curve of the self-constructed dataset, comparing YOLOv5 with SFER-YOLOv5.

Table 1. Experimental results on self-constructed dataset.

Expression	YOLOv5		SFER-YOLOv5	
	P	*mAP*@0.5	*P*	*mAP*@0.5
happy	0.816	0.853	0.851	0.887
sad	0.845	0.867	0.889	0.909
neutral	0.842	0.857	0.879	0.871
all	0.834	0.859	**0.873**	**0.889**

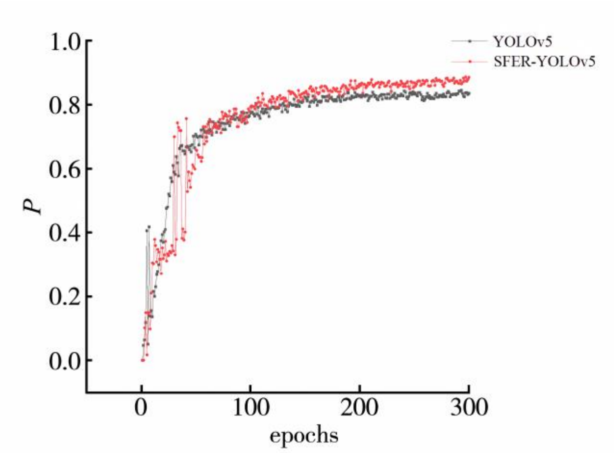

Fig.2 P curve on the self-constructed dataset.

Table 2 presents the experimental results on the Fer-2013 dataset. Fig. 3 illustrates the *P* curve of the Fer-2013 dataset.

Table 2. Experimental results on FER-2013 dataset.

Expression	YOLOv5		SFER-YOLOv5	
	P	*mAP*@0.5	*P*	*mAP*@0.5
happy	0.863	0.918	0.876	0.911
sad	0.636	0.594	0.669	0.626
neutral	0.694	0.799	0.722	0.515
all	0.731	0.770	**0.756**	**0.784**

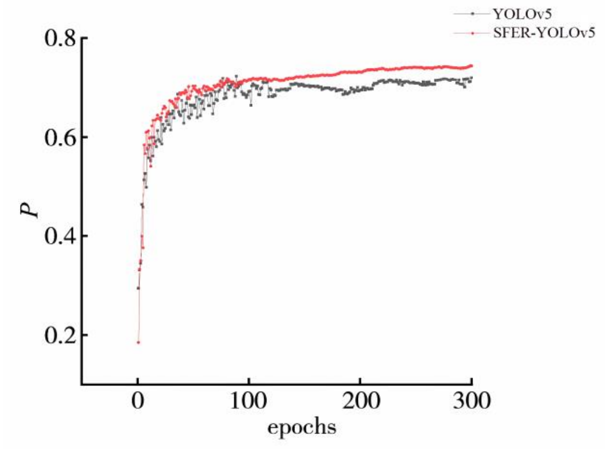

Fig.3 P curve on the FER-2013 dataset.

5. Conclusion

This study addresses the challenges of low real-time detection rates and poor timeliness in recognizing students' facial expressions in complex classroom environments while simultaneously focusing on mitigating and preventing cyberbullying incidents. Our *SFER-YOLOv5* model improves the non-maximum suppression (NMS) by replacing it with Soft NMS, enhances the feature extraction capability by adding the Coordinate Attention (CA) mechanism module, and improves the representation ability of target boxes by using EIoU instead of CIoU. Our experimental results are highly encouraging, demonstrating a 3.9% increase in detection accuracy (P) and a 3.0% increase in mAP@0.5 on the self-constructed dataset. Beyond facial expressions, future research will integrate poses and gestures for a comprehensive analysis of online classroom dynamics. This enhances our cyberbullying prevention by understanding students' engagement and emotions more comprehensively.

References

1. J. Bao, X. Tao, Y. Zhou, "An Emotion Recognition Method Based on Eye Movement and Audiovisual Features in MOOC Learning Environment", IEEE Transactions on Computational Social Systems, 2022.
2. P. Ekman, W.V. Friesen, "Constants across cultures in the face and emotion", Journal of personality and social psychology, 17.2 (1971), 124.
3. N. Shakee, R. K. Dwivedi, "A survey on detection of cyberbullying in social media using machine learning techniques", Intelligent Communication Technologies and Virtual Mobile Networks: Proceedings of ICICV 2022, Singapore: Springer Nature, pp. 323-340, 2022.

4. A. Bozyiğit, S. Utku, E. Nasibov, "Cyberbullying detection: Utilizing social media features", Expert Systems with Applications, 179: 115001, 2021.
5. R. Kumar, A. Bhat, "A study of machine learning-based models for detection, control, and mitigation of cyberbullying in online social media", International Journal of Information Security, 21(6), 1409-1431, 2022.
6. G.W. Giumetti, R.M. Kowalski, "Cyberbullying via social media and well-being", Current Opinion in Psychology, 45, 101314, 2022.
7. X. Zhu, et al, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios", Proceedings of the IEEE/CVF international conference on computer vision, pp. 2778-2788, 2021.

Authors Introduction

Mr. Shuai Wang

He is currently pursuing the Ph.D. degree in Computer Science programme with the Institute of Computer Science and Digital Innovation, UCSI University, Kuala Lumpur, Malaysia. He is also working in the Department of Physics and Electronic Engineering, Yuncheng University, Yuncheng, Shanxi, China. His current research interests include computer vision, machine learning and multi-modal learning.

Dr. Abdul Samad Bin Shibghatullah

He received the Ph.D. degree in computer science from the Brunel University of Uxbridge, United Kingdom. He is currently an Associate Professor at Institute of Computer Science & Digital Innovation, UCSI University, Kuala Lumpur, Malaysia. His current research interests include optimization, modelling and scheduling, Robotics.

Dr. Javid Iqbal Thirupattur

He works as an Assistant professor at UCSI University, Malaysia. He holds a Ph.D. degree in Information and Communication Technology from the National Energy University Malaysia. His research interests include Multimedia, Augmented Reality, and Virtual Reality.

Dr. Keoy Kay Hooi

He received the Ph.D. degree at Sheffield Hallam University, United Kingdom in 2006. He is currently an Associate Professor, UCSI University, Kuala Lumpur, Malaysia. His current research interests include machine Learning, block chain technology & management information system.