SUGISAKA MASANORI

Journal of Robotics, Networking and Artificial Life Vol. 10(4); March (2024), pp. 357–361
ISSN (Online): 2352-6386; ISSN (Print): 2405-9021
https://alife-robotics.org/jrnal.html

Research Article

YOLOv5 Based Student Engagement and Emotional States Detection in E-Classes

Shuai Wang[1,2], Abdul Samad Shibghatullah[3], Kay Hooi Keoy[4], Javid Iqbal[5],

[1]*Institute of Computer Science and Digital Innovation, UCSI University, Cheras, Kuala Lumpur, Malaysia*
[2]*Department of Physics and Electronic Engineering, Yuncheng University, Yuncheng, Shanxi, China*
[3]*College of Computing & Informatics, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia*
[4]*Graduate Business School, UCSI University, Cheras, Kuala Lumpur, Malaysia*
[5]*Department of Computing and Information Systems, Sunway University, Petaling Jaya, Selangor, Malaysia*

ARTICLE INFO

Article History

Received 28 November 2023
Accepted 18 July 2024

Keywords

Cyberbullying detection
Emotion recognition
Deep learning
E-learning
YOLO

ABSTRACT

The rapid expansion of E-learning environments has highlighted the critical issue of cyberbullying within digital classrooms. This study introduces a novel approach for early detection of cyberbullying by analyzing student engagement and emotional states in real time. Our SER-YOLO model fuses an advanced You Only Look Once version 5 (YOLOv5) with a Student Emotion Recognition system, enriched by sophisticated methodological improvements. It features Soft NMS to refine the Non-Maximum Suppression (NMS) process, embeds the Channel Attention (CA) module to augment the network's backbone, and employs Enhanced Intersection over Union (EIOU) for bounding box regression. This method proactively detects changes in student engagement and emotional states, providing an effective mechanism for the early detection and management of cyberbullying in E-learning environments.

1. Introduction

In the era of the big data revolution, propelled by machine learning (ML) and artificial intelligence (AI), and with the rapid growth of virtual learning, online social media platforms have become an integral part of modern educational frameworks. The unexpected advent of the COVID-19 in 2020 has significantly altered traditional approaches to learning, work, and daily life. The previously unchallenged model of face-to-face instruction is now confronted with significant obstacles. Many academic institutions have turned to online instruction as a practical alternative. However, this shift to digital learning environments has brought forth a range of challenges, with cyberbullying being particularly prominent. Cyberbullying, alternatively termed as online harassment or digital aggression, involves the intentional use of electronic means to cause harm through aggression, mockery, threats, and other malicious behaviors. The virtual classroom (or online classroom), a new arena for both teachers and students, poses potential risks of cyberbullying that can severely impact the emotional and psychological well-being of its participants, as well as hinder the educational experience. Moreover, cyberbullying has been linked to a decline in student engagement. The emotional state is also vulnerable to fluctuations due to the stress and anxiety induced by the presence of cyberbullying, leading to a less conducive atmosphere for educational growth and exchange.

In online learning classrooms, communication between students and teachers is facilitated solely through screens, limiting the educator's assessment of comprehension to the students' facial expressions and emotional states. Analyzing these expressions can assist teachers in better understanding the students' levels of engagement and in making timely adjustments to their teaching methods [1].

Corresponding author E-mail: 1002268166@ucsiuniversity.edu.my

In 1971, Paul Ekman et al. [2] carried out comprehensive studies of human facial expressions, distinguishing six fundamental emotional expressions: happiness, surprise, fear, sadness, disgust, and anger. These basic emotions are pivotal for interpreting students' reactions in the classroom. Positive emotions like happiness and surprise indicate active engagement and willingness to learn, while negative emotions such as sadness, anger, fear, and disgust suggest a lack of interest or inattention. Neutral expressions may suggest moderate engagement. Thus, monitoring students' facial expressions in an online setting is crucial for effective teaching and learning interactions.

Amidst these challenges, educational entities and digital platforms are increasingly focusing on the effective identification and mitigation of cyberbullying within the virtual classroom milieu. The advent of facial expression recognition (FER) technology, with its prospects for emotion analytics and affective state monitoring, has garnered considerable interest. Despite the burgeoning interest, the extant research leveraging the YOLO family for facial expression recognition remains sparse. This paper introduces an innovative approach to facial expression recognition (FER) predicated on an enhanced YOLOv5 architecture, intended for the prompt identification and amelioration of cyberbullying in the online classroom setting.

Our paper's key contributions are: (1) Enhanced Soft-NMS: Replaces the standard NMS, enabling the capture of subtle expressions crucial for detecting student emotional states. (2) CA Mechanism: Improves the recognition of complex behaviors in scenarios involving multiple students. (3) Loss Function Enhancement: Utilizes EIOU and Focal Loss to effectively mitigate the challenges posed by sample imbalance.

2. Related work

2.1. *Cyberbullying in E-classes*

The rise of online learning, accelerated by the COVID-19 pandemic, has brought both opportunities and challenges to the academic community. A significant challenge is cyberbullying, which includes different types of online harassment within virtual classrooms, such as verbal abuse, intimidation, misinformation, and exclusion from classroom activities [3], [4]. These behaviors have profound impacts on students' academic performance, mental health, and social relationships [5].

Addressing cyberbullying in virtual classrooms is critical. Research has explored various methods to identify and prevent such behavior, including analyzing behavioral patterns, parsing linguistic content, examining social network structures, applying machine learning algorithms, and detecting emotional cues. These approaches aim to anticipate and mitigate bullying incidents in digital learning environments.

Additionally, numerous studies highlight the importance of educational initiatives and improving cyberbullying awareness among faculty and students. These efforts are essential for creating a safe and supportive online learning environment [6]. Implementing these strategies not only protects against the harmful effects of cyberbullying but also reinforces the commitment to maintaining the dignity and well-being of all participants in virtual education.

2.2. *YOLOv5*

In May 2020, Ultralytics LLC introduced YOLOv5 [7], an efficient convolutional neural network designed for swift image processing, capable of remarkable inference speeds, reaching up to 140 frames per second (FPS), thus satisfying the stringent real-time demands for video analysis. This model, crafted with Python, is a testament to the progress in lightweight deep learning architectures. YOLOv5 is available in four scalable variants: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, all sharing a fundamental design philosophy. These variants are distinguished by parameters such as width_multiple and depth_multiple parameters, which regulate the network's capacity and complexity, respectively. Consequently, each variant offers a unique balance between model size and accuracy, allowing for tailored selection based on the computational and accuracy requirements of specific applications. Among these, the Yolov5s variant, with its reduced depth and parameter count, stands out as an optimal candidate for real-time applications such as facial expression analysis, where rapid inference is paramount.

By optimizing network structures and training strategies, YOLOv5 enables rapid and accurate facial expression recognition in videos, crucial for assessing student engagement and emotional states in classrooms. Research shows that integrating YOLOv5 with emotion classification models can enhance the analysis of students' emotional states, aiding teachers in understanding learning experiences and adjusting instructional methods to improve quality.

For cyberbullying detection, YOLOv5 can be combined with technologies like natural language processing (NLP) to identify and prevent bullying behaviors in online environments. Analyzing students' online behaviors and language expressions allows YOLOv5 to promptly detect potential bullying incidents and facilitate timely interventions.

In this research, we employ YOLOv5s version 6.0 as a foundation for further advancements. We present the

SER-YOLO model, a groundbreaking methodology that harnesses the sophisticated YOLOv5 framework to accurately detect cyberbullying incidents via facial expression analysis in digital classrooms and online learning spaces. This model represents a significant step towards creating a safer and more emotionally intelligent virtual educational experience.

3. Methodology

The current literature on facial expression recognition leveraging the YOLOv5 framework, while not extensive, presents a compelling opportunity for monitoring student engagement and emotional states in online educational spaces. Our contribution to this field builds upon the existing YOLOv5s-6.0 model by introducing a series of enhancements that significantly elevate its performance on experimental datasets.

3.1. *Data collection and processing*

To precisely discern the facial expressions associated with varying degrees of students' comprehension—namely, mastery, confusion, and non-mastery—we conducted a questionnaire-based survey. This study aimed to explore the correlation between these levels of understanding and the corresponding facial expressions exhibited. The survey spanned higher vocational colleges and universities, yielding a total of 520 responses (287 male, 233 female). We focused on seven fundamental emotional expressions: anger, surprise, disgust, happiness, fear, sadness, and neutral.

In this research, we utilized both the Fer2013 dataset and a proprietary collection of student facial expressions captured within an educational setting. The Fer2013 dataset [8], known to include non-facial and incorrectly labeled images, underwent a meticulous process of selection, cleansing, and filtration to enhance the precision of our experiments, culminating in a refined set of 11,000 images. Concurrently, our in-house dataset was curated by surveilling classroom videos, compiling 668 images that were subsequently labeled after extracting relevant frames from the recorded sessions.

3.2. *YOLOv5 Enhancements*

3.2.1. *Optimized NMS Mechanism*

The methodologies of NMS and its variant, Soft NMS, rely on the confidence scores derived from classification predictions, with higher scores signifying greater localization precision. For the advancement of detection accuracy in this research, we have implemented an enhanced version of Soft NMS, which surpasses the efficacy of the traditional NMS approach. The ultimate calculation is expressed in Eq. (1).

$$s_i = s_i e^{-\frac{iou(B,b_i)^2)}{\sigma}}, \forall b_i \notin D \quad (1)$$

This method ensures that some high-scoring bounding boxes are preserved as accurate detections in subsequent computations.

3.2.2. *Coordinate attention module*

In our research, we introduce the Coordinate Attention (CA) mechanism [9], a novel approach that integrates spatial coordinates with channel attention. Our experiments substantiate that incorporating the CA module into the core network framework enhances detection accuracy. This improvement is particularly noticeable in classroom environments with multiple students, especially those seated in distant rows from the camera's view. These advancements are credited to the CA module (see Fig. 1), which strengthens channel characteristics in the feature map, enabling the capture of more precise and influential data. The network architecture with the CA module integrated into the backbone is illustrated in Fig. 2.

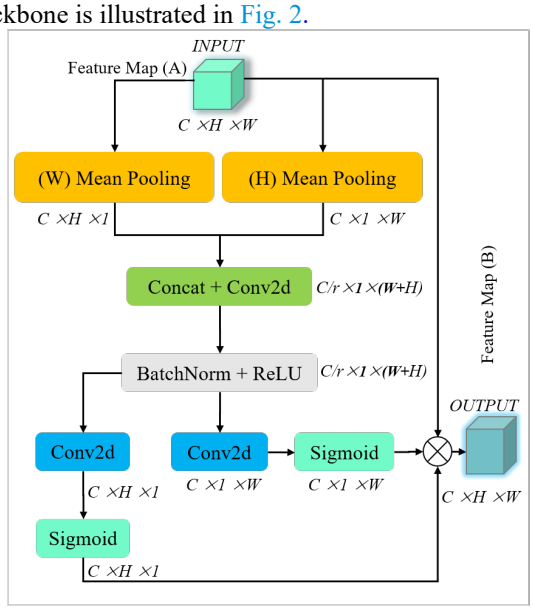

Fig. 1 The mechanism of CA.

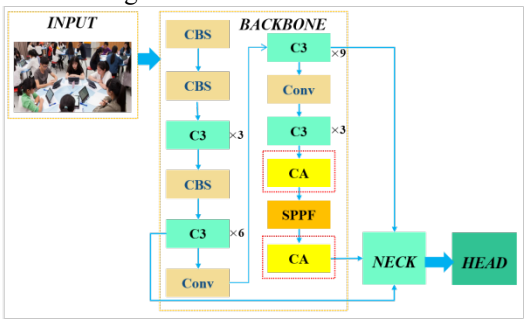

Fig. 2 The enhanced network of YOLOv5s-6.0.

3.2.3. *Refined EIOU Loss*

Our approach enhances the original loss function by integrating the Efficient-IOU (EIOU) loss function to calculate height and width-related loss values, thereby improving upon conventional aspect ratio estimation based on CIOU. Furthermore, we have incorporated Focal Loss to mitigate class imbalance, thereby achieving more balanced optimization across diverse samples. The formulas for EIOU and Focal Loss [10] are provided in Eq. (2) and (3), respectively.

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{(w^c)^2+(h^c)^2} + \frac{\rho^2(w,w^{gt})}{(w^c)^2} + \frac{\rho^2(h,h^{gt})}{(h^c)^2} \quad (2)$$

$$L_{Focal}(p_t) = -\alpha(1-p_t)\beta\, ln(p_t) \quad (3)$$

where, α is a factor serves to address sample imbalance, and β adjusts the weights between hard and easy samples.

4. Experimental results

Our study's experimental setup featured a 64-bit Windows 10 OS, powered by an 11th Gen Intel Core i5-11400H processor and an NVIDIA RTX 3050 graphics card. We utilized PyTorch 1.8 for deep learning, within a Python 3.8 version environment. Table 1 details the hyperparameter settings used during the model training process. The training process spanned 300 epochs.

Fig. 3 illustrates the overall SER-YOLO method. The model training utilized pre-trained weights (yolov5s.pt) and a configuration file (SER.YAML). Logs and weight files were saved throughout the training process.

For comprehensive assessment, we conducted extensive experiments on the self-constructed dataset and Fer2013. Table 2 compares the performance of the YOLOv5 and SFER-YOLO models on experimental datasets [11]. SFER-YOLO significantly outperforms the baseline YOLOv5. Precision improved from 83.4% to 87.3%, a 3.9% increase, while mAP@0.5 rose from 85.9% to 88.9% after 150 epochs, marking a 3.0% gain. These results confirm SFER-YOLO's enhanced effectiveness in detecting and tracking student engagement.

Table 1 The hyperparameter settings of model training.

Hyper-parameters	Value
resolution	640x640
optimizer	Adam
lr0	0.01
lrf	0.1
weight_decay	0.0005
epoch	300
batch_size	16
IOU threshold	0.5

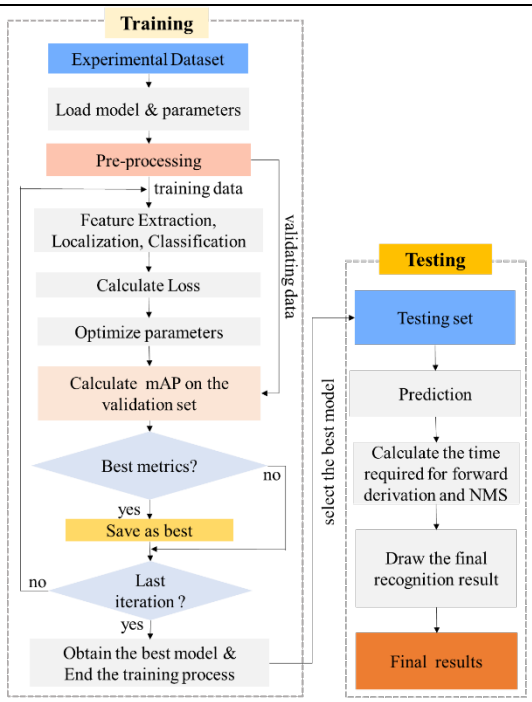

Fig. 3 The overall process of SER-YOLO model

Table 2 Experimental results on our self-constructed and FER-2013 dataset

Emotion	Self-constructed Dataset				FER-2013			
	YOLOv5		*SER-YOLO*		*YOLOv5*		*SER-YOLO*	
	P	*mAP*@0.5	*P*	*mAP*@0.5	*P*	*mAP*@0.5	*P*	*mAP*@0.5
happy	0.816	0.853	0.851	0.887	0.863	0.918	0.876	0.911
sad	0.845	0.867	0.889	0.909	0.636	0.594	0.669	0.626
neutral	0.842	0.857	0.879	0.871	0.694	0.799	0.722	0.515
all	0.834	0.859	**0.873**	**0.889**	0.731	0.770	**0.756**	**0.784**

5. Conclusion

This research addresses the pivotal issues of inadequate real-time detection capabilities and the lack of timeliness in identifying students' facial expressions within the intricate backdrop of e-learning settings, concurrently focusing on the mitigation of cyberbullying incidents. Our SER-YOLO model introduces an enhanced soft NMS to replace the conventional method, enriches feature extraction through an optimized Coordinate Attention (CA) mechanism, and refines bounding box representation by adopting the EIoU_loss function instead of CIoU_loss. The empirical findings indicate a notable 3.9% enhancement in precision (*P*) and a 3.0% increase in *mAP@0.5* on our self-constructed dataset. Subsequent studies will broaden the scope to encompass not only facial expressions but also body postures and gestures, thereby providing a more comprehensive understanding of the dynamics within online classrooms and strengthening cyberbullying prevention through deeper insights into student participation and emotional states.

Acknowledgment

This study was supported by the 2023 Applied Research Project of Yuncheng University (Project no. YY-202312).

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. M. Khairy, T.M. Mahmoud, A. Omar, T. Abd El-Hafeez, "Comparative performance of ensemble machine learning for Arabic cyberbullying and offensive language detection", Language Resources and Evaluation, 58(2), 695-712, 2024.
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. K. Zhou, et al., "Evaluation of BFRP strengthening and repairing effects on concrete beams using DIC and YOLO-v5 object detection algorithm", Construction and Building Materials, 411, 134594,2024.
8. P. Giannopoulos, I. Perikos, I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013", Advances in hybridization of intelligent methods: Models, systems and applications, 1-16, 2018.
9. Y. Li, M. Zhang, C. Zhang, H. Liang, P. Li, "YOLO-CCS: Vehicle Detection Algorithm Based on Coordinate Attention Mechanism", Digital Signal Processing, 104632, 2024.
10. C. Liu, et al., "Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism", Neural Networks, 170, 276-284, 2024.
11. S. Wang, A. S. Shibghatullah, J. Iqbal, K. H. Keoy, "Online Classroom Student Engagement Analysis using Enhanced YOLOv5", ICAROB, 974-978, 2024.

Authors Introduction

Mr. Shuai Wang

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

Dr. Abdul Samad Bin Shibghatullah

He earned his Ph.D. in Computer Science from Brunel University, Uxbridge, UK. He is now an Associate Professor Universiti Tenaga Nasional, Malaysia. His research interests focus on optimization, modeling, and scheduling.

Dr. Keoy Kay Hooi

He earned his Ph.D. from Sheffield Hallam University, UK, in 2006. He is currently an Associate Professor at the Centre for Business Informatics and Industrial Management, Graduate Business School (GBS), UCSI University, Malaysia. His research interests include machine learning and block chain.

Dr. Javid Iqbal Thirupattur

He is a Senior Lecturer at Sunway University, Malaysia, holding a Ph.D. in Information and Communication Technology from the National Energy University, Malaysia. His research interests are in multimedia, augmented reality, and virtual reality.