www.ijtes.net

Face Anonymization in Intelligent Experiment Education

Jiangyi Cui
School of Information Science and Technology, Fudan University, China

Ruijiao Li
Fudan Academy for Engineering and Technology, Fudan University, China

Qiushu Chen
School of Information Science and Technology, Fudan University, China

Libin Liu
Shanghai Xiding AI Research Center Co., Ltd, China

Xuan Zhao
School of Information Science and Technology, Fudan University, China

Kai Liu
Shanghai Xiding AI Research Center Co., Ltd, China

Huiliang Shang
Shanghai Xiding AI Research Center Co., Ltd, China

To cite this article:

Cui, J., Li, R. Chen, Q., Liu, L., Zhao, X., Liu, K., & Shang, H. (2025). Face anonymization in intelligent experiment education. *International Journal of Technology in Education and Science (IJTES), 9*(3), 434-449. https://doi.org/10.46328/ijtes.641

IJTES International Journal of Technology in Education and Science
2025, Vol. 9, No. 3, 434-449 https://doi.org/10.46328/ijtes.641

Face Anonymization in Intelligent Experiment Education

Jiangyi Cui, Ruijiao Li, Qiushu Chen, Libin Liu, Xuan Zhao, Kai Liu, Huiliang Shang

Article Info

Article History

Received:
27 January 2025
Accepted:
20 June 2025

Keywords

AIED
Anonymization protection
YOLOv8
Face segmentation
Occlusion

Abstract

Face anonymization in intelligent experimental education is crucial for privacy protection. This paper presents a novel, real-time face blurring system for smart experimental settings. Our key contributions include: 1) customized YOLOv8 (Multi-Scale Feature Fusion YOLOv8) algorithm achieving 96% accuracy at 22.67 fps for 1080p video. 2) An annotation dilation preprocessing method, Contour-Adaptive Occlusion Refinement (CAOR), to address instrument occlusion issues for training. 3) A specialized dataset of 51 experimental videos with dense annotations. Our system tackles the unique challenge of preserving experimental details while anonymizing faces. We introduce two metrics, Sensitivity of Blur Accuracy (SOBA) and Over Blurred Rate (OBR), to evaluate performance. Our work demonstrates robustness across physics, biology, and chemistry experiments, maintaining a low mis-blur rate of 0.02 for instruments.

Introduction

Intelligent education systems have revolutionized learning environments, particularly in scientific experiments. AIED (Artificial Intelligence in Education) applications and platforms（Agca, et al., 2025; Thinger, et al, 2024; Aldahdouh, et al,2024; Ozdere, et al, 2025;） such as Knewton (Ali, et al., 2022), Dreambox Learning (Gellen, et al., 2024), and other adaptive learning systems have gained prominence (Jing, et al., 2023; Badali, et al., 2022). These systems collect and store student learning data, leveraging big data and machine learning methods to analyze and provide feedback on learning outcomes (Bittencour, et al., 2024), thereby creating personalized learning plans (Yang, et al., 2024). In scientific experiment education, computers can create virtual experimental environments using VR or AR technologies (Kubincová, et al., 2023; Yücel, et al., 2024; Putra, et al., 2024; Zaturrahmi, et al., 2020), offering students more realistic laboratory scenarios. Additionally, intelligent experiment platforms have been introduced into classrooms, where devices capture students' experiment processes and results, analyzing their learning and operational skills (Wenyang, et al, 2023). Beyond classroom learning, applications like Edmodo (Nurrohma, et al., 2021) and Google Classroom (Iftakhar, et al., 2016; Trisnawati, et al., 2024) provide online educational platforms, enabling students to interact with teachers and peers both within and outside their classes (Torun, et al., 2025).

Privacy concerns have emerged as a critical issue in AIED. AIED typically processes the collected data using methods such as data anonymization and de-identification, where sensitive personal information is deleted or obscured. The stored information is encrypted, and access to data is restricted to verified personnel, ensuring

secure data control through facial recognition, fingerprint verification, or account authentication. Face anonymization is one of the key issues for privacy protection. Current facial privacy protection methods, including multi-object tracking with blur and pixilation (Zhou, et al., 2020; Shang, et al., 2021), k-same clustering algorithms (Pan, et al., 2019; Yang, et al., 2024), and GAN-based approaches for virtual face generation (Wang, et al., 2023; Kumar, et al., 2024) perform well in generic face anonymization contexts, but are not tailed for the unique need of smart experimental education.

Scientific experiments present unique face anonymization challenges in intelligent education. Video recordings capture students' faces, operational techniques, and instrument usage simultaneously, which demands a delicate balance: students' facial identities must be anonymized for both personal privacy and fair assessment, while instruments and manipulation actions must remain clearly visible for accurate evaluation. Furthermore, the complexity lies in scientific apparatus pose additional challenges, as they often include fine structures (e.g., wires) which are easily overlooked or confused with background noise, or transparent objects (e.g., beakers, test tubes) which lacks clear boundaries and exhibit variable optical effects. Traditional anonymization methods struggle with this balance. Blurring or pixelation techniques risk obscuring crucial experimental details. Our challenge is to develop an algorithm that fulfills unique combination of requirements:

1. Accurately detect and anonymize faces in real-time.
2. Precisely segment and preserve instrument visibility, including slender structures in cluttered environments and transparent objects with complex optical properties (refraction, reflection, transparency).
3. Maintain clear visibility of students' operational techniques.

We propose an innovative facial blurring algorithm that selectively preserves frontal experimental instruments while anonymizing faces. Our method is built on YOLOv8, a versatile model for image detection and segmentation. It incorporates a simple yet effective annotation pre-processing algorithm to improve separation between small frontal experimental objects and facial pixels requiring blurring. To evaluate our method objectively, we introduce two new metrics: SOBA (Sensitivity of Blur Accuracy) and OBR (Over Blurred Rate). This approach achieves an optimal balance between accuracy and speed, ensuring precise facial anonymization with efficient processing. As is shown in figure 1, the system performs fine anonymization on the primary experimenter's face, accurately segmenting the facial contours and instruments for privacy protection. For non-primary experimenters, the system applies coarse processing without detailing facial contours.

Figure 1. The Effect of the Face Anonymization System on Experimental Video Frames

Our main contributions are as follows:

- We proposed a comprehensive face anonymization framework for intelligent experimental education systems built upon an ad-hoc customized YOLOv8 architecture: Multi-Scale Feature Fusion YOLOv8 (MSFF-YOLOv8). This framework achieves real-time facial blurring at 22.67 fps (1920 * 1080 pixels) with 96% accuracy during experiments, demonstrating robustness across physics, biology, and chemistry experimental environments. It provides targeted solutions when faces are obscured by fine or transparent instruments, ensuring facial anonymization while clearly presenting operational techniques. Notably, it maintains a low mis-blur rate of 0.02 for experimental instruments, offering a secure and reliable privacy protection approach.
- We integrated a dilation module, Contour-Adaptive Occlusion Refinement (CAOR) for experimental instrument masks within the data processing segment. This module performs outward dilation from the central point of the experimental instrument mask, ensuring that pixels around the instrument's outline are not misidentified. This guarantees the clear visibility of both the experimental instruments and the operations conducted. This module reduces the system's OPR (over blur rate) by approximately 10%.
- To support our research, we created a specialized dataset. We collected 51 videos of experimenters from different experiments, both frontal and side views, using intelligent education devices. These videos were manually annotated with dense labels (7590 face detection and segmentation labels, 5240 instrument labels) to create a dataset suitable for face detection and segmentation. Additionally, we processed various poses of the experimenters within the frames (1530 poses) and created a face tracking dataset categorized by the experimenters.

Literature Review

Current face anonymization techniques can be broadly classified into three categories: multi-object tracking (MOT)-based methods, clustering-based methods, and generative adversarial network (GAN)-based methods.
MOT-based methods are commonly used for detecting and tracking faces in video sequences (Albiero, et al., 2021; Liu, et al., 2023; Yang, et al., 2024). In the initial frame, face detection algorithms identify and localize faces by generating bounding boxes. Tracking algorithms then follow the trajectory of these faces across subsequent frames (Wang, et al., 2020; Conklin, et al., 2016). For anonymization, the pixels within the detected bounding boxes are either blurred or pixelated. Clustering-based methods, such as the K-Same algorithm (Hocine, et al., 2024; Meden, et al., 2018), aim to anonymize faces by replacing them with the average of the K most similar faces in the dataset. This process involves preprocessing the facial images (e.g., aligning them by rotation) and calculating the distances between them to identify similar faces. The original face is then substituted with the average of its K nearest neighbors.

GAN-based methods utilize generative adversarial networks to regenerate anonymized faces (Rai, et al., 2024). For example, Deep Privacy (Hukkelås, et al., 2019), employs conditional GANs (CGANs) to generate new faces by extracting key facial features (e.g., eyes and eyebrows) through face detection and feature recognition. These features are then used as input for generative models, such as diffusion networks or adversarial generators, to synthesize new faces. These methods are effective in removing privacy-sensitive information while producing

high-quality anonymized face images. However, under certain challenging conditions, such as non-traditional poses or significant face occlusion, the quality of the generated images may not always meet expectations. It is a novel way that suits the tasks with high quality of face fidelity but no time requirement. Because it requires significant computational resources and is commonly used in image processing applications. For instance, minusface takes 69ms to convert images for Xp protection, which already represents the lightweight version of this type of method (Peng, et al., 2024; Mi, et al., 2024).

Common methods for facial processing include skin color thresholding and deep learning-based approaches. The skin color thresholding method involves converting an RGB image to the YCrCb color space and setting a threshold range to distinguish skin pixels from other background pixels. This method struggles to effectively differentiate between hand and face pixels. In recent years, there have been numerous segmentation models emerging in the field of deep learning. Models like YOLOv8 from the YOLO series (Terven, et al., 2023) have integrated segmentation tasks alongside detection and classification tasks. Lightweight networks such as U-Net (Siddique, et al., 2021) and Mask-RCNN (He, et al., 2017), tailored for medical image segmentation, have also shown promising performance in various tasks. The trend of large models is gradually rising, particularly in the segmentation field where models like SAM and Edge SAM are leading efforts to extract object contours.

These models segment all objects in images or use human-computer interaction to segment individual or multiple objects, with humans providing points or bounding boxes as prompts. This approach can also be deployed locally. Challenges of face anonymization in scientific experiment scenes are mainly two points. First, the intelligent experiment platform input is video, which has high requirements for the simplicity and real-time performance of the algorithm. There is a large amount of data in the video, and complex algorithms will have higher requirements for the performance of the equipment and affect the processing speed. Secondly, the instruments and students' techniques in scientific experiments are crucial for teachers' observations and grading. These details must not be obscured when applying face blurring, and the algorithm must accurately distinguish and preserve objects such as instruments and hand movements that may be in front of the face, ensuring that these essential elements remain clear and unaltered. This requires the algorithm to pay attention not only to the blur of the face range but also to the accurate of the occlusion range, which has not been proposed in the field of segmentation. In order to meet this requirement, we innovatively propose dilation data processing and segmentation model optimization.

Methods

Overview

Our real-time face anonymization system for intelligent experimental education integrates multiple components to ensure privacy protection while preserving experimental details. We trained modified YOLOv8 models for face detection and segmentation to localize and extract face pixels. This is combined with a BOTSort multi-object tracker for face tracking across video frames, followed by Gaussian blurring for face anonymization.

In experimental videos, faces are frequently occluded by students' hands or instruments such as glass jars and tubes. These occlusions can impede face segmentation and lead to misclassification of obstructing objects as face

pixels. Consequently, this results in unintended blurring of instruments and hand motions, which are essential elements for evaluation in intelligent scientific experiment systems.

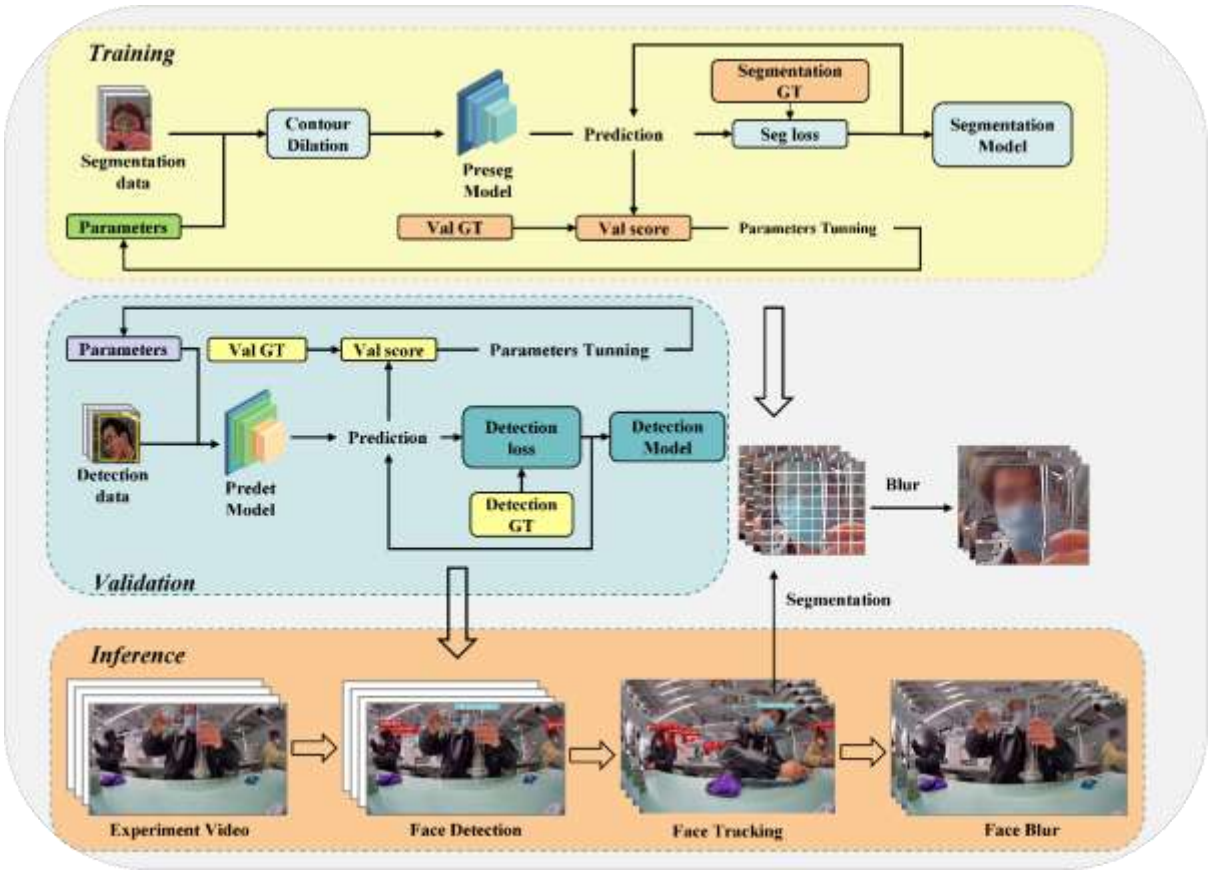

Figure 2. The Pipeline of our Face Anonymization Algorithm

To mitigate this issue, we developed a segmentation annotation pre-processing algorithm based on contour dilation. This algorithm reduces face pixel misclassification and subsequent edge mis-blurring, ensuring that experimental actions and instruments remain visible. We also modified the YOLOv8 backbone and enhanced the shortcut low-level feature extractor for both segmentation and object detection heads, improving the model's ability to capture fine details.

Our pipeline, comprising training, validation, and inference stages, is illustrated in Figure 2. The training stage involves data processing, application of our algorithm, and model training. During validation, we utilize multiple scoring functions to evaluate model performance and adjust parameters. In the inference stage, we employ the trained models for face anonymization while maintaining experimental observation integrity.

Algorithm Design

Our algorithm addresses the challenge of protecting student privacy in intelligent educational equipment while maintaining the visibility of experimental operations. We developed a face detection and segmentation system that identifies and tracks faces in scientific experiment videos, localizes operators, and applies occlusion and face

segmentation blur to the primary operator. To optimize processing speed, the algorithm distinguishes between primary and non-primary operators in the detection and tracking module. For non-primary operators, we directly blur their bounding boxes, thereby reducing the computational load on the segmentation module.

The core of our segmentation and blurring system is built upon an improved YOLOv8 architecture. This modified structure retains convolutional features from lower layers (e.g., P2) to enhance small object detection and precision. We introduced a new feature fusion method to strengthen the interaction between shallow and deep feature layers, resulting in improved multi-scale detection performance. A significant challenge in experimental settings is the occlusion of faces by instruments or students' hands. To address this, we implemented a dilation data processing method. This technique extends the edges of occlusions towards the face, effectively preventing unintended blurring of critical experimental objects.

Contour-Adaptive Occlusion Refinement (CAOR)

During face segmentation training and inference, down sampling and other operations reduce image resolution, sacrificing detailed features and making it difficult to accurately distinguish between the contours of faces and instruments. Additionally, the movement of operators or instruments causes motion blur, leading to unclear or trailing edges of objects in single-frame images.

To this end, CAOR is specifically designed to improve the quality of training data for face segmentation models, with a crucial focus on preserving apparatus and instruments as non-face areas. This method is applied exclusively during the training phase to enhance the model's ability to accurately distinguish between faces and experimental equipment.

Algorithm Contour-Adaptive Occlusion Refinement
Input:
face_mask: face binary mask of the main operator
occlusion_mask: binary mask of the occlusion in front of operator’s face
Output:
processed_mask: face mask after the processing of occlusion mask’s edge dilation

```
Function FindCenterOfFace(face_mask):
    total_x, total_y, total_pixels = 0, 0, 0
    for each pixel (i, j) in binary_mask do
        if face_mask[i][j] == 1 then
            total_x += j
            total_y += i
            total_pixels += 1
    center_x = total_x / total_pixels
    center_y = total_y / total_pixels
    return center_x, center_y
```

```
intersection_mask = AND (face_mask, occlusion_mask)
intersection_contours = FindContours(intersection_mask)
# Calculate Dilation scale factor
face_center = FindCenterOfFace(face_mask)
scale_factor = CalScaleFactor (face_mask, occlusion_mask)
scale_factor = ApplyGaussianBlur(scale_factor)
# Calculate Dilation vector and add new point to the original contour
for each contour in intersection_contours do
  for each point (x, y) in contour do
    vector_to_center = (face_center.x - x, face_center. y - y)
    vector_length = sqrt (vector_to_center.x^2 + vector_to_center. y^2)
    vector_length = max (vector_length, epsilon)
    normalized_vector = (vector_to_center.x / vector_length, vector_to_center. y / vector_length)
    expanded_direction = (normalized_vector.x * scale_factor, normalized_vector. y * scale_factor)
    expansion_distance = scale_factor
    new_x = x + expanded_direction.x * expansion_distance / vector_length
    new_y = y + expanded_direction. y * expansion_distance / vector_length
    add (new_x, new_y) to scaled_contour
  end for
add scaled_contour to scaled_contours
end for
Processes_mask = CrprimaryeateMaskFromContours(scaled_contours)
```

When occlusion occurs, we use segment annotations as the initial contour for faces and occluding objects, with the overlapping boundary between the occlusion and the face as the boundary to be expanded, and the direction pointing towards the face center as the expansion direction. Since the movement of objects varies across frames, we adaptively determine the dilation coefficient d, by assessing the recognizability of overlapping targets in the recognition area without compromising face anonymization. This allows the overlapping edges to expand towards the face, effectively addressing the issue of misblurring at the edges of occlusions caused by motion blur and feature extraction.

Multi-Scale Feature Fusion YOLOv8 (MSFF-YOLOv8)

MSFF-YOLOv8 addresses the potential loss of fine-grained features during downsampling in the original YOLOv8 architecture. This enhanced model introduces a novel feature fusion method to improve the preservation and utilization of detailed information across different scales(figure 3, the red arrows indicate the modified feature transmission and fusion). In the Neck component, MSFF-YOLOv8 incorporates additional feature connection paths, specifically between shallow feature layers (such as P2) and deeper layers. This is achieved through increased Upsample and Concat operations, fostering stronger interactions between multi-scale features. The frequency of up-sampling and down-sampling operations in the feature pyramid is also increased, optimizing

information flow between shallow features (P2, P3) and deep features (P4, P5). These modifications significantly enhance the model's capability to detect small objects.

The Head component of MSFF-YOLOv8 introduces a semantic segmentation branch, integrating additional ConvModule and CSP2Conv modules after each feature layer's output. This design enables simultaneous object detection and semantic segmentation, effectively implementing multi-task learning within the model. By fusing multi-level features and concatenating them layer by layer, MSFF-YOLOv8 preserves both fine-grained information and multi-scale contextual data. This comprehensive approach aims to enhance the model's overall accuracy and robustness, particularly in complex scenarios. The improvements are designed to offer a more refined solution for detection and segmentation tasks, with a focus on multi-scale and small object detection performance.

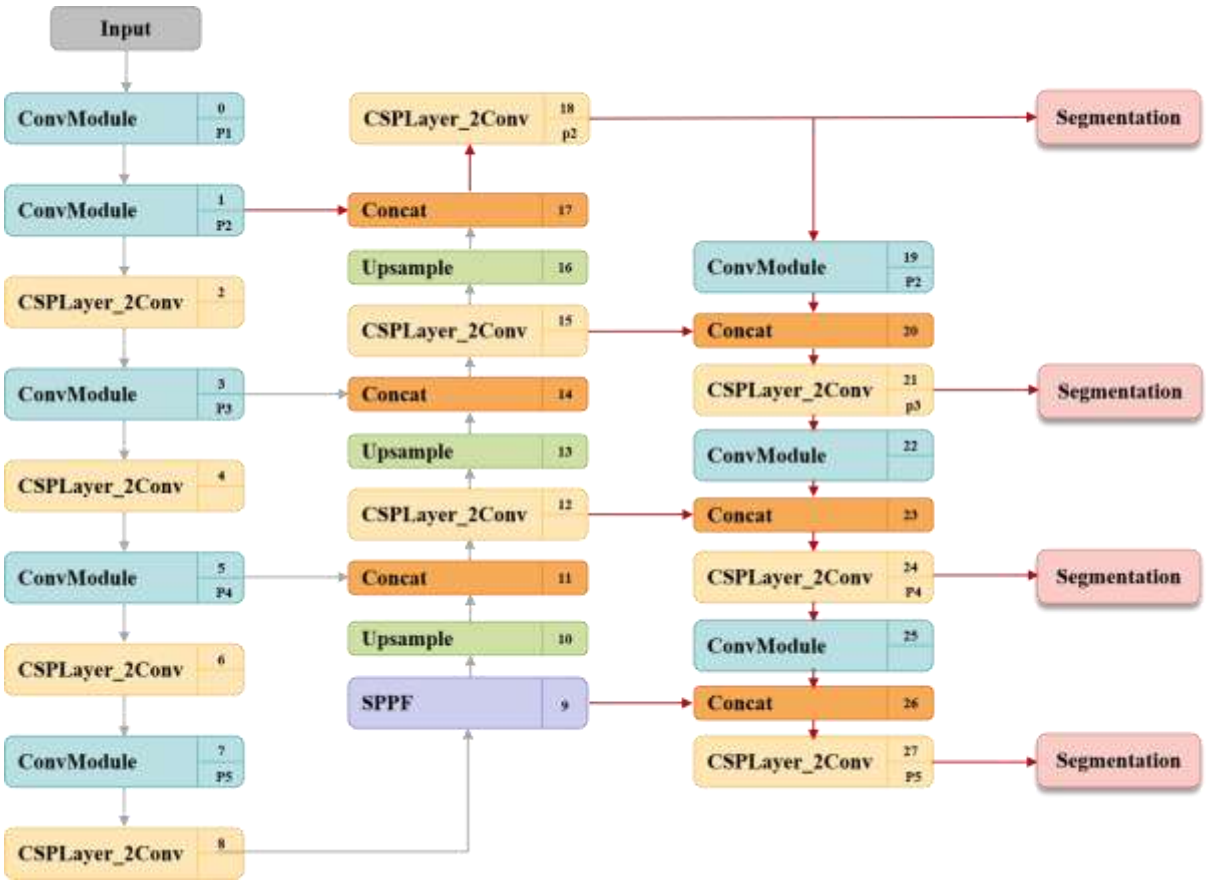

Figure 3. Diagram of MSFF-YOLOv8 Model Structure

Experiment and Discussion

We created a dataset for face detection, segmentation, and tracking tasks, focusing on scientific experiment videos, which captures student faces in real time. To evaluate our approach, we conducted experiments using both the MSFF-YOLOv8 network and the original YOLOv8 network, as well as SAM (Segment Anything Model) and Edge SAM. The comparisons were based on three key metrics: frames per second (FPS) to evaluate processing speed, SOBA to reflect segmentation accuracy, and OPR, a metric specifically designed for experimental

scenarios, to measure the retention of occluded objects. This comprehensive evaluation allowed us to effectively analyze the overall effectiveness of the framework tailored for experimental scenarios, as well as the impact of CAOR method and YOLOv8 network optimizations.

Dataset

We collected 51 videos of various experiments such as "balance measurement", "light refraction", "pinhole imaging" and "density measurement of substances" using intelligent educational equipment and manually annotated face detection boxes and related pixels as benchmarks. We ultimately selected 17 experiment videos in which students were frequently occluded as occluded dataset. Frames were extracted every 10 seconds from these videos, resulting in 1080*1920 RGB images, and a dataset of 1049 images which combines both seldom-occluded and frequent occluded dataset. Each image was annotated with face bounding boxes and face pixels to create both face detection and face segmentation datasets. When the same face was occluded by instruments and divided into several parts, each part was labeled as a face. Additionally, we organized the different postures of each operator to create an object tracking dataset.

Metrics

In this study, we aim to evaluate the performance of the algorithm under various conditions using three key metrics: Frame Rate (FPS), Sensitive Object Blur Accuracy (SOBA), and Over Blur Rate (OBR) [15]. The FPS metric is used to quantify the algorithm's processing speed by measuring the number of frames processed per second, serving as a primary indicator of real-time efficiency. Higher FPS values reflect the algorithm's ability to handle large volumes of data efficiently without compromising performance. SOBA, on the other hand, measures the algorithm's ability to accurately handle face blurring within video frames. This metric specifically evaluates how well the algorithm maintains object boundaries during dynamic scenes, with higher values indicating more accurate blurring management. OBR focuses on the mis blurring of occluded objects, assessing the algorithm's effectiveness in segmenting occluded regions. A lower OBR value is desirable as it reflects improved performance in distinguishing and accurately segmenting occluded faces or objects, thus minimizing the detrimental effects of blur on recognition accuracy. Together, these metrics offer a comprehensive evaluation of the algorithm's performance in terms of speed, accuracy in dynamic conditions, and its ability to manage occlusions.

$$SOBA = 1 - \frac{\sum_t \left(m_t + f_{pt} + mm_t\right)}{\sum_t \left(g_t\right)} \quad (1)$$

Where g_t is total pixels, mm_t are pixels of faces that were not detected, m_t is detected pixels that were incorrectly identified as non-facial, f_{pt} is Pixels incorrectly identified as facial in frame t.

$$OBR = \frac{\sum_t f_{ot}}{\sum_t c_t} \quad (2)$$

Where c_t are all matched pixels, f_{ot} are pixels that are mistakenly marked as human faces.

SOBA and OBR is intuitive explained in figure 4, (a) shows the case where the algorithm performs well, serving as the baseline. SOBA represents the accuracy of detecting facial pixels—the higher the value, the better. In (b), the two images have a lower SOBA compared to (a), indicating fewer detected facial pixels and lower accuracy. OBR represents the rate of erroneous blurring of instruments and operations—the lower the value, the better. A lower OBR means fewer pixels of instruments are misidentified as facial pixels. In (c), there is a high SOBA and a high OBR, indicating a higher accuracy in facial detection compared to (b), but also an erroneous blurring of instruments and operations.

Figure 4. Comparison of Effects of Different SOBA and OBR

Evaluation of Algorithm Performance

Due to the lack of publicly available datasets, we used our proposed dataset. We trained and tested two models, YOLOv8 and MSFF-YOLOv8, on datasets that were either processed with or without CAOR, to evaluate the impact of CAOR on model performance. Additionally, we compared these models with other segmentation models such as SAM and Edge-SAM. We randomly selected 20 segments of 5-minute video clips and compared the models' blurring accuracy (SOBA) and frame rate (FPS) against manually annotated data. We trained the face detection model on YOLOv8 and set the point in bounding box as a prompt for face segmentation based on SAM and Edge-SAM.

In particular, we tested some heavily occluded video segments to assess the models' segmentation and blurring performance under complex occlusion scenarios, especially focusing on cases of mis blurring (OBR). Figure 5 compares results of YOLOv8 model and MSFF-YOLOv8 model with varying parameters. The rightmost column indicates the model used for each row. Column (a) shows two original experimental images. Column (b) and (c) have an image size of 640, with column (b) without CAOR and column (c) processed with CAOR. Columns (d) and (e) have an image size of 1920, with column (d) without CAOR and column (e) with CAOR.

SAM is one of the most popular methods for segmentation tasks and was the first method we attempted in our preliminary experiments. However, due to the lack of semantic information, the YOLO series is more suitable for

this task. In the experiments, we compared it with YOLOv8 model. We trained multiple models for testing, comparing the performance of YOLOv8m and MSFF-YOLOv8, as well as the effect of incorporating dataset CAOR processing. MSFF-YOLOv8 is the new model network structure proposed by us. We randomly selected 20 segments of 5-minute videos from our captured footage for facial blur testing, comparing accuracy and inference speed against manually annotated data.

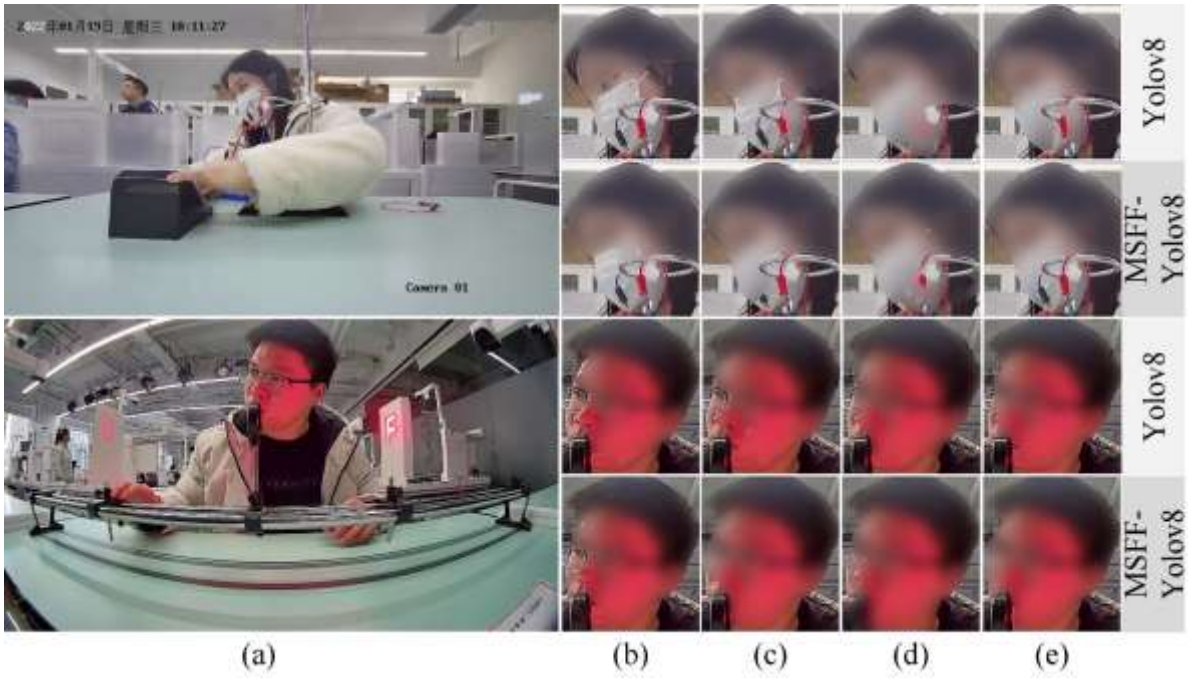

Figure 5. Comparison of different YOLOv8 Model with Varying Parameters.

Table 1. Evaluation of Different Network Structures and Data Processing Methods

Parameters			**Appearance**	
Model	**Network Structure**	**Data processing**	**SOBA（↑）**	**Fps（↑）**
SAM	original	original	0. 6852 (0.0729↓)	1.05 (13.44↓)
Edge-SAM	original	original	0. 5907 (0.1674↓)	5.46 (9.03↓)
YOLOv8	original	original	0. 7581 (Baseline)	14.49 (Baseline)
YOLOv8+CAOR	original	dilation	0. 8598 (0.1017 ↑)	14.49 (-)
MSFF-YOLOv8	improved	original	0. 8582 (0.1001↑)	22.67 (8.18↑)
MSFF-YOLOv8 +CAOR	improved	dilation	**0. 9337 (0.1756↑)**	**22.67 (8.18↑)**

In Table 1, ↑ indicates that a higher value for the metric is desirable and the increase in SOBA and FPS values is considered positive. A value followed by ↑ indicates how much the metric has im-proved compared to the baseline. Conversely, a value followed by ↓ indicates how much the metric has decreased compared to the

baseline. As is shown in the table, the processing speed of SAM and edge-SAM when handling experimental videos frame by frame is unsuitable for high-real-time video processing scenarios. Additionally, their accuracy in experimental segmentation is slightly lower compared to YOLOv8. MSFF-YOLOv8 demonstrates higher real-time performance compared to the original version. With default parameters, its processing speed reaches 14.49 FPS, which is an improvement of 8.18 FPS over YOLOv8. This meets the real-time processing requirements of face privacy protection systems in educational scenarios.

In terms of accuracy, the model trained on datasets preprocessed with CAOR achieves a higher accuracy for blurred images compared to the model trained on unprocessed datasets. In actual experimental scenarios, 70% of video frames may be in an unobstructed state. We manually selected some heavily occluded videos for testing to evaluate the algorithm's segmentation and blurring performance under instrument or procedural obstruction, particularly assessing cases of erroneous obstruction of instruments versus procedures.

The improved YOLOv8 model demonstrated great performance in both original and heavily obstructed videos shown in Table 2 (original full dataset) and table 3 (occluded dataset), especially after the targeted data dilation processing, which significantly reduced the OBR indicator. The MSFF-YOLOv8 model generally outperforms the original YOLOv8. In the MSFF-YOLOv8 model, directly transmitting features from P2 and P3 layers to the head for prediction actually lowered accuracy at low resolutions. Based on comprehensive experimental results, setting image size to 1920 after modifying the model achieved the highest accuracy and minimized erroneous obstruction rates. However, under the default input size, both YOLOv8 and the proposed model show suboptimal performance compared to when the input size is adjusted. Therefore, in practical applications, it is crucial to reasonably select and fine-tune the input size to achieve optimal results.

Table 2. Comparison of Different Parameters on Full Dataset

Parameters		**Appearance**	
Model	**Image size**	**SOBA（↑）**	**OBR（↓）**
SAM	640	0. 6852 (0.0729↓)	0.1038 (0.0187↓)
Edge-SAM	640	0. 5907 (0.1674↓)	0.1268 (0.0043↑)
YOLOv8	640	0. 7581 (Baseline)	0.1225 (Baseline)
YOLOv8 + CAOR	640	0. 8598 (0.1017↑)	0.0928 (0.0297↓)
YOLOv8	1920	0. 9484 (0.1903↑)	0.1037 (0.0188↓)
YOLOv8 + CAOR	1920	0. 9547 (0.1966↑)	0.0174 (0.1051↓)
MSFF-YOLOv8	640	0. 8582 (0.1001↑)	0.1158 (0.0067↓)
MSFF-YOLOv8 + CAOR	640	0. 9337 (0.1756↑)	0.0457 (0.0768↓)
MSFF-YOLOv8	1920	0. 9507 (0.1926↑)	0.0792 (0.0433↓)
MSFF-YOLOv8 + CAOR	1920	**0. 9602 (0.2021↑)**	**0.0159 (0.1066↓)**

Table 3. Comparison of Different Parameters on Occluded Dataset

Parameters		**Appearance**	
Model	**Image size**	**SOBA（↑）**	**OBR（↓）**
SAM	640	0. 5907 (0.1402↓)	0.1268 (0.0181↓)
Edge-SAM	640	0. 4267 (0.3042↓)	0.1686 (0.0237↑)
YOLOv8	640	0. 7309 (Baseline)	0.1449 (Baseline)
YOLOv8 + CAOR	640	0. 8237 (0.0928↑)	0. 0934 (0.0515↓)
YOLOv8	1920	0. 9061 (0.1752↑)	0. 1369 (0.0080↓)
YOLOv8 + CAOR	1920	0. 9124 (0.1815↑)	0. 0242 (0.1207↓)
MSFF-YOLOv8	640	0. 8074 (0.0765↑)	0. 1288 (0.0161↓)
MSFF-YOLOv8 + CAOR	640	0. 8727 (0.1418↑)	0. 0480 (0.0969↓)
MSFF-YOLOv8	1920	0. 9207 (0.1898↑)	0. 0834 (0.0615↓)
MSFF-YOLOv8 + CAOR	1920	**0. 9221 (0.1912↑)**	**0. 0167 (0.1282↓)**

Conclusion

This paper introduces a novel face anonymization system tailored for scientific experimental education scenarios within AIED. The system leverages an enhanced YOLOv8 model (MSFF-YOLOv8), incorporating improved low-level feature retention and fusion techniques, alongside strategic dilation processing of training data, Contour-Adaptive Occlusion Refinement (CAOR). This approach achieved 96.02% accuracy and maintained a real-time 1080p video processing speed of 22.67 fps. Our system successfully achieves face anonymization while preserving crucial experimental details. This balance is particularly vital in intelligent scientific experimental education scenarios, where maintaining the visibility of apparatus is as important as protecting privacy.

While our implementation of CAOR dilation processing during the data preparation phase significantly mitigates edge pixel misclassification issues, we acknowledge that further research is necessary to achieve precise edge classification during real-time inference. This presents an exciting avenue for future work, potentially leading to even more robust and accurate face anonymization in dynamic educational environments.

References

Adhikari, Y. N., & Rana, K. (2024). Sustainability Challenges of Universities' Online Learning Practices. *International Journal of Technology in Education and Science, 8*(3), 430-446.

Albiero, V., Chen, X., Yin, X., Pang, G., & Hassner, T. (2021). img2pose: Face alignment and detection via 6dof, face pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7617-7627).

Aldahdouh, T. Z., Al-Masri, N., Abou-Dagga, S., & AlDahdouh, A. (2024, January). *Development of online teaching expertise in fragile and conflict-affected contexts. In Frontiers in Education* (Vol. 8, p. 1242285). Frontiers Media SA.

Ali, N., Ullah, S., & Khan, D. (2022). Interactive laboratories for science education: A subjective study and systematic literature review. *Multimodal technologies and interaction,* 6(10), 85.

Alkan, A. (2024). The Role of Artificial Intelligence in the Education of Students with Special Needs. *International Journal of Technology in Education and Science, 8*(4), 542-557.

Badali, M., Hatami, J., Banihashem, S. K., Rahimi, E., Noroozi, O., & Eslami, Z. (2022). The role of motivation in MOOCs' retention rates: a systematic literature review. *Research and Practice in Technology Enhanced Learning, 17*(1), 5.

Bittencourt, I. I., Chalco, G., Santos, J., Fernandes, S., Silva, J., Batista, N., ... & Isotani, S. (2024). *Positive artificial intelligence in education (P-AIED): A roadmap. International Journal of Artificial Intelligence in Education, 34*(3), 732-792.

Ceylan, B., & Karakus, M. A. (2024). Development of an Artificial Intelligence-Based Mobile Application Platform: Evaluation of Prospective Science Teachers' Project on Creating Virtual Plant Collections in Terms of Plant Blindness and Knowledge. *International Journal of Technology in Education and Science, 8*(4), 668-688.

Conklin, T. A. (2016). *Knewton* (An adaptive learning platform available at https://www. knewton. com/).

Gellen, S., Wicker, K., Ainsworth, S., Morris, S., & Lewin, C. (2024). *Using the DreamBox Reading Plus adaptive literacy intervention to improve reading attainment, a two-armed cluster randomised trial evaluation protocol.*

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *In Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Hocine, L., & Leghima-aissat, A. (2024). *Enhancing B2B Marketing with AI: A New Era of Service Efficiency Cases of Salesforce Education Cloud, Knewton, and Coursera*, 1365-1380.

Hu, W., Liu, K., Liu, L., & Shang, H. (2023). Hu, W., Liu, K., Liu, L., & Shang, H. (2023). A Spatial-Temporal Transformer based Framework for Human Pose Assessment and Correction in Education Scenarios. arXiv preprint arXiv:2311.00401.. arXiv preprint arXiv:2311.00401.

Hukkelås, H., Mester, R., & Lindseth, F. (2019, October). Deepprivacy: A generative adversarial network for face anonymization. In *International symposium on visual computing* (pp. 565-578). Cham: Springer International Publishing.

Iftakhar, S. (2016). Google classroom: what works and how. *Journal of education and social sciences, 3*(1), 12-18.

Jing, Y., Zhao, L., Zhu, K., Wang, H., Wang, C., & Xia, Q. (2023). *Research landscape of adaptive learning in education: A bibliometric study on research publications from 2000 to 2022. Sustainability, 15*(4), 3115.

Kubincová, Z., Hao, T., Capuano, N., Temperini, M., Ge, S., Mu, Y., ... & Yang, J. *Emerging Technologies for Education.*

Kumar, L., & Singh, D. K. (2024). Diversified realistic face image generation GAN for human subjects in multimedia content creation. *Computer Animation and Virtual Worlds, 35*(2), e2232.

Liu, Z., Wang, X., Wang, C., Liu, W., & Bai, X. *SparseTrack: Multi-Object Tracking by Performing Scene*

Decomposition Based on Pseudo-Depth. arXiv 2023. arXiv preprint arXiv:2306.05238.

Meden, B., Emeršič, Ž., Štruc, V., & Peer, P. (2018). k-Same-Net: k-Anonymity with generative deep neural networks for face deidentification. *Entropy, 20*(1), 60.

Mi, Y., Zhong, Z., Huang, Y., Ji, J., Xu, J., Wang, J., ... & Zhou, S. (2024). Privacy-preserving face recognition using trainable feature subtraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 297-307).

Nurrohma, R. I., & Adistana, G. A. Y. P. (2021). Penerapan Model Pembelajaran Problem Based Learning dengan Media E-Learning Melalui Aplikasi Edmodo pada Mekanika Teknik. *Edukatif: Jurnal Ilmu Pendidikan, 3*(4), 1199-1209.

Pan, Y. L., Haung, M. J., Ding, K. T., Wu, J. L., & Jang, J. S. (2019, September). K-same-siamese-gan: K-same algorithm with generative adversarial network for facial image de-identification with hyperparameter tuning and mixed precision training. In 2019 16th *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-8). IEEE.

Peng, T., Li, M., Chen, F., Xu, Y., Xie, Y., Sun, Y., & Zhang, D. (2024). ISFB-GAN: Interpretable semantic face beautification with generative adversarial network. *Expert Systems with Applications, 236*, 121131.

Putra, A. S., & Zainul, R. (2024). Serious Games in Science Education: A Review of Virtual Laboratory Development for Indicator of Acid-Base Solution Concepts. *Chemistry Smart, 3*(1), 1-8.

Rai, A., Gupta, H., Pandey, A., Carrasco, F. V., Takagi, S. J., Aubel, A., ... & De la Torre, F. (2024). Towards realistic generative 3d face models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3738-3748).

Shang, Y., & Wong, H. C. (2021). Automatic portrait image pixelization. *Computers & Graphics, 95*, 47-59.

Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE access, 9,* 82031-82057.

Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction, 5*(4), 1680-1716.

Thinger, C., Verma, A. S., Thakore, A. K., Purohit, P., Kalyanwat, B., & Sikarwal, P. (2024). *AIED-Artificial Intelligence in Education World-Opportunities, Challenges, and Future Research. Challenges, and Future Research* (November 14, 2024).

Trisnawati, N. F., Setyo, A. A., Sundari, S., & Warlatu, A. (2024). Improving Mathematical Reasoning Skills Through an Open-Ended Approach Assisted by Google Classroom and Google Meet. *Mosharafa: Jurnal Pendidikan Matematika, 13(*2), 489-502.

Wang, T., Zhang, Y., Zhao, R., Wen, W., & Lan, R. (2023). Identifiable face privacy protection via virtual identity transformation. *IEEE Signal Processing Letters, 30*, 773-777.

Wang, Z., Zheng, L., Liu, Y., Li, Y., & Wang, S. (2020, August). *Towards real-time multi-object tracking. In European conference on computer vision* (pp. 107-122). Cham: Springer International Publishing.

Yang, F., Li, K., & Li, R. (2024). AI in language education: Enhancing learners' speaking awareness through AI-supported training. *International Journal of Information and Education Technology, 14*(6), 828-833.

Yang, M., Han, G., Yan, B., Zhang, W., Qi, J., Lu, H., & Wang, D. (2024, March). Hybrid-sort: Weak cues matter for online multi-object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol.

38, No. 7, pp. 6504-6512).

Yang, Z., Luo, L., Gu, Y., & Ren, F. (2024, July). K-Face Net: A Two-Stage Framework for Balanced Feature Space in Facial Expression Recognition. In 2024 *IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). IEEE.

Yücel, F., Ünal, H. S., Surer, E., & Huvaj, N. (2024). *A Modular Serious Game Development Framework for Virtual Laboratory Courses*. IEEE Transactions on Learning Technologies.

Zaturrahmi, Z., Festiyed, F., & Ellizar, E. (2020). The utilization of virtual laboratory in learning: A meta-analysis. *Indonesian Journal of Science and Mathematics Education, 3(*2), 228-236.

Zhou, J., Pun, C. M., & Tong, Y. (2020, October). Privacy-sensitive objects pixelation for live video streaming. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 3025-3033).

Author Information

Jiangyi Cui
https://orcid.org/0009-0005-4385-2814
School of Information Science and Technology, Fudan University
China

Ruijiao Li
https://orcid.org/0000-0002-3191-3828
Fudan Academy for Engineering and Technology, Fudan University
China

Qiushu Chen
https://orcid.org/0000-0002-5679-2188
School of Information Science and Technology, Fudan University
China

Kai Liu
https://orcid.org/0000-0002-1203-019X
Shanghai Xiding AI Research Center Co., Ltd
China

Libin Liu
https://orcid.org/0000-0001-9900-5933
Shanghai Xiding AI Research Center Co., Ltd
China

Huiliang Shang
https://orcid.org/0000-0002-9553-1589
School of Information Science and Technology, Fudan University
China

Xuan Zhao
https://orcid.org/0000-0002-2443-3500
Yiwu Research Institute of Fudan University
China
Contact e-mail: *zhaox@fudan.edu.cn*