Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

PRIVA: Selective Face Blurring Video App Using YOLOv8-Face and MobileFaceNet

Muhammad Satrio[1)], Mohammad Nasucha[2)*]
[1)]Department of Informatics, Universitas Pembangunan Jaya, Indonesia
[2)]Center for Urban Studies, Universitas Pembangunan Jaya, Indonesia
*Corresponding Author
muhammad.satrio@student.upj.ac.id, mohammad.nasucha@upj.ac.id

Submitted :May 25, 2026 | **Accepted** : June 11, 2026 | **Published** : July 5, 2026

Abstract: The increasing use of vlog videos on social media creates privacy risks because third-party faces are often unintentionally recorded and distributed without consent. Existing face blurring approaches generally apply uniform anonymization to all detected faces and do not provide an identity-selective mechanism that keeps the content creator visible while blurring other individuals. This study develops PRIVA, a desktop-based selective face blurring application that runs locally without an external AI server. The proposed pipeline integrates YOLOv8n-Face-960 for face detection, MobileFaceNet for face recognition using 512-dimensional embeddings, and Deep SORT for maintaining identity consistency across video frames. Face enrollment is performed through guided multi-pose webcam capture, while video evaluation is conducted on extracted YOLO analysis frames from five real vlog-like test videos. YOLOv8n-Face-960 achieved an overall detection precision of 95.02%, recall of 89.32%, and F1-score of 92.09%. The baseline comparison showed that YOLOv8n-Face-960 achieved a higher mean detection F1-score than MTCNN, while MobileFaceNet provided a smaller and faster recognition model than FaceNet for CPU-based local inference. For correctly detected face instances, PRIVA achieved a system precision of 99.45%, recall of 98.70%, F1-score of 99.08%, and accuracy of 98.50% in determining whether faces should be blurred or kept visible. Processing performance testing showed an average analysis speed of 4.83 FPS, average export speed of 70.05 FPS, and average processing ratio of approximately 2.40 times the original video duration. These results indicate that PRIVA can support practical local identity-selective face blurring for video privacy protection, although detection robustness remains important under low-light, crowded, distant, or partially occluded face conditions.

Keywords: Deep SORT, face detection, face recognition, MobileFaceNet, ONNX Runtime, selective face blurring

INTRODUCTION

The widespread use of social media platforms has changed how people create and distribute visual content online. One of the most popular formats is the video blog, or vlog, because it captures daily activities in a spontaneous and personal way. According to DataReportal, Indonesia had 143 million social media user identities in January 2025, representing 50.2% of the total population, with YouTube reaching 67.3% of internet users in the country (Kemp, 2025). Vlog and influencer video content is also widely consumed in Indonesia, where 33.3% of internet users aged 16 to 64 watch this type of content every week (We Are Social & Hootsuite, 2023). This condition encourages creators to record videos in public or semi-public spaces, where the appearance of other individuals in the frame is often unavoidable.

The presence of third-party faces in video content creates privacy risks because the human face is a biometric identifier that can be used to recognize, track, or misuse a person's identity. Facial images distributed online can be indexed by facial search engines, manipulated into deepfake content (Mirsky & Lee, 2021), or exploited for doxing and identity theft (Wang et al., 2025). In Indonesia, Law Number 27 of 2022 on Personal Data Protection classifies biometric data, including facial images, as sensitive personal data that requires explicit consent before processing or publication (Republik Indonesia, 2022). Similar protection is also established by the General Data Protection Regulation in the European Union (European Parliament & Council of the European Union, 2016). In

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

addition to legal obligations, social media platforms also enforce privacy policies that may result in video removal or loss of monetization when identifiable faces are published without consent. Therefore, content creators need a practical way to protect the identities of individuals who are unintentionally captured in their videos.

Manually blurring the faces of strangers is a time-consuming process because video editing often requires frame-by-frame inspection and adjustment. Standard digital video commonly runs at 30 to 60 frames per second, meaning that one minute of footage may contain 1,800 to 3,600 frames that need to be checked. Existing video editing software generally provides manual keyframe-based blur tools, while automated face anonymization systems usually apply the same blur operation to all detected faces. Previous studies on automated face de-identification and face blurring have shown that deep learning can be used to detect and obscure faces automatically (Khan et al., 2025; Laishram, Shaheryar, et al., 2024; Plaud & Lisani, 2024). However, these approaches generally do not provide an identity-selective mechanism that keeps the content creator visible while blurring non-consenting individuals. Platform-based tools also remain limited because they are tied to specific services and cannot be applied directly to a reusable video file before distribution across multiple platforms.

Based on these limitations, this study develops PRIVA, a desktop-based application for automatic identity-selective face blurring in video. The application is built using Rust and the Tauri framework so that the processing workflow can run locally on the user's device without relying on an external AI server (Kosasih W. & Engel, 2026). In the final implementation, PRIVA integrates YOLOv8n-Face-960 for face detection, MobileFaceNet for face recognition, and Deep SORT for maintaining identity consistency across video frames (Chen et al., 2018; Wojke et al., 2017). The system allows users to enroll their own face, analyze imported videos, review detected face tracks, and export a processed video in which non-enrolled faces are blurred while selected enrolled faces remain visible. Through this approach, PRIVA is expected to provide a practical local tool for helping content creators reduce the risk of publishing identifiable third-party faces in social media videos.

LITERATURE REVIEW

Research on automated facial privacy has shown that deep learning can be used to detect, blur, or de-identify faces in images and videos. (Wang et al., 2025) categorized facial privacy preservation methods into several paradigms, while (Khan et al., 2025) reviewed person de-identification methods and showed that many existing approaches still apply privacy protection uniformly to detected faces or persons. (Laishram, Lee, et al., 2024) proposed face de-identification using caricature generation, and (Plaud & Lisani, 2024) demonstrated automatic face blurring in videos using deep learning. Although these studies contribute to facial privacy protection, they generally do not provide an identity-selective workflow that preserves selected enrolled identities while blurring non-enrolled individuals in the same video. To clarify the novelty of this study, Table 1 compares previous facial privacy studies with PRIVA based on several functional aspects required for identity-selective face blurring in video.

Table 1. Comparison of previous facial privacy studies and PRIVA.

Study	Main Focus	Limitation Compared with PRIVA
(Plaud & Lisani, 2024)	Automatic face blurring.	Does not support identity-selective blur.
(Khan et al., 2025)	Person de-identification review.	Survey-level study, not a local video editing workflow.
(Laishram, Lee, et al., 2024)	Face de-identification using caricature.	Image-based, not selective video blurring.
(Wang et al., 2025)	Facial privacy survey.	Survey-level study, not an implemented desktop app.
PRIVA	Identity-selective face blurring in video.	Local desktop workflow with detection, recognition, tracking, review, and export.

The comparison shows that PRIVA differs from previous facial privacy studies by focusing on identity-selective face blurring for vlog videos in a local desktop workflow. This task requires reliable face detection, face recognition, and temporal identity consistency. YOLO-based detectors are suitable for efficient object detection because they perform detection in a single pass (Redmon et al., 2016), while YOLOv8 improves localization using an anchor-free detection head (Jocher et al., 2023). MobileFaceNet is used because it provides efficient face embedding extraction for local recognition compared with heavier recognition models such as FaceNet (Chen et al., 2018; Schroff et al., 2015). Deep SORT is used to maintain identity consistency across video frames by

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16273 p-ISSN : 2541-044X

combining motion prediction and appearance-based association (Wojke et al., 2017). The application is deployed using Tauri to support a lightweight local desktop workflow without server reliance (Kosasih W. & Engel, 2026). Therefore, the contribution of PRIVA lies in integrating YOLOv8n-Face-960, MobileFaceNet, Deep SORT, user review, and local video export into a single identity-selective face blurring application.

METHOD

This section explains the development and evaluation process of PRIVA, including the research design, AI model integration, system architecture, application implementation, and testing procedure using real video data.

Research Design

This study employs a quantitative approach within a prototyping research method. The quantitative approach is used because PRIVA is evaluated using measurable outputs, including detection results, embedding similarity, tracking results, blur decisions, and processing time. The prototyping method is used because the system was developed through iterative improvements from earlier face detection and recognition configurations into the final identity-selective face blurring pipeline. System performance is evaluated using detection coverage, selective face blurring performance, processing performance, baseline component comparison, and blackbox functional testing. The detection evaluation measures the ability of YOLOv8n-Face-960 to detect visible faces in extracted analysis frames, while the selective blur evaluation measures whether the system correctly keeps enrolled faces visible and blurs non-enrolled faces. The baseline comparison evaluates MTCNN against YOLOv8n-Face-960 for detection and FaceNet against MobileFaceNet for recognition efficiency. The evaluation results are reported using per-video and aggregated metrics.

Evolutionary Development

PRIVA was developed through three prototype phases. Phase 1 used MTCNN and FaceNet in a Python-based prototype, but MTCNN showed limitations in detecting non-frontal and difficult face poses. Phase 2 replaced MTCNN with YOLOv8-Face to improve detection coverage and added SORT for tracking, but FaceNet remained relatively large for lightweight local deployment. Phase 3 became the final implementation by integrating YOLOv8n-Face-960 for detection, MobileFaceNet for recognition, and Deep SORT for identity tracking. MobileFaceNet was selected because the ONNX model used in this study is smaller than the FaceNet baseline, with 12.99 MB compared with 89.61 MB, while remaining suitable for local face recognition. The application was also migrated to Rust and Tauri 2.0 so that all AI inference processes can run locally on the CPU through ONNX Runtime without relying on an external AI server.

Data Collection and Face Enrollment

The identity reference data used by PRIVA is collected through guided webcam-based enrollment. Each enrolled individual follows five pose directions: frontal, left, right, upward, and downward. The system captures 20 samples per person, consisting of 6 frontal, 4 left, 4 right, 3 upward, and 3 downward samples. For each sample, PRIVA performs face detection, preprocessing, and embedding extraction using YOLOv8n-Face-960 and MobileFaceNet. The resulting 512-dimensional L2-normalized embedding is stored in the face library for identity matching. Raw enrollment frames are not used as a training dataset because PRIVA relies on pre-trained models and does not perform additional model training.

Preprocessing Pipeline

PRIVA applies the same preprocessing pipeline during enrollment and video analysis. Each input image or YOLO analysis frame is letterboxed and resized to 960x960 pixels for YOLOv8n-Face-960, which produces face bounding boxes and five facial landmarks. Each detected face is then prepared as a 112x112 input for MobileFaceNet. If valid landmarks are available, PRIVA applies an Umeyama similarity transformation to align the face to a canonical five-point template. Otherwise, it uses a padded bounding-box crop resized to 112x112. The resulting RGB face crop is normalized from [0, 255] to [-1, 1] using Equation (1).

$$x_{norm} = \frac{x - 127.5}{127.5} \qquad (1)$$

The normalized face crop is passed into MobileFaceNet to produce a 512-dimensional embedding vector. The embedding is then L2-normalized so that identity matching can be performed using cosine similarity.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

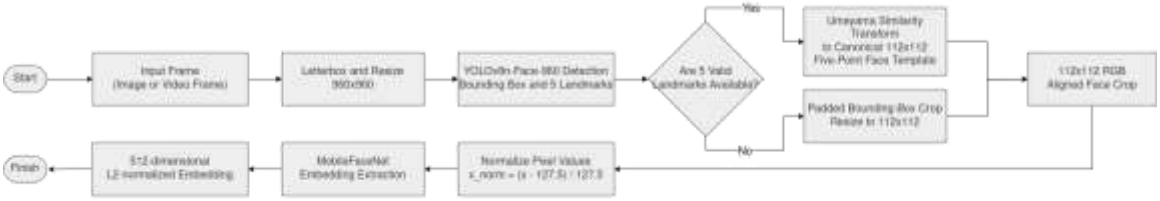


Fig. 2 Face preprocessing pipeline from input frame to 512-dimensional L2-normalized embedding, including landmark-based alignment and fallback cropping.

Model Validation

Before system-level testing, the detection and recognition models were validated to ensure that their outputs could be used in the PRIVA pipeline. YOLOv8n-Face-960 was validated by comparing predicted bounding boxes with manually annotated ground truth boxes on extracted analysis frames. A detection was considered correct when the predicted box matched a ground truth face with an Intersection over Union value of at least 0.5, as shown in Equation (2).

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \qquad (2)$$

MobileFaceNet was validated by comparing detected face embeddings with stored enrollment embeddings. Each detected face was converted into a 512-dimensional L2-normalized embedding, and identity similarity was calculated using cosine similarity, as shown in Equation (3).

$$similarity(A, B) = \frac{A \cdot B}{||A||||B||} \qquad (3)$$

The predicted identity was determined from the nearest enrolled embedding above the matching threshold of 0.40. The validation ensured that the detector could locate visible faces and that the recognition model could compare detected faces against enrolled identities before the complete tracking and blur decision workflow was evaluated.

System Design and Architecture

The PRIVA pipeline is implemented as a local desktop application with a Rust and Tauri 2.0 backend and a React, TypeScript, and Vite frontend. AI inference runs locally on the CPU through ONNX Runtime, using YOLOv8n-Face-960 with 960x960 input for face detection and MobileFaceNet with 112x112 input for face recognition. FFmpeg and FFprobe are used for video decoding, probing, and export, while frontend-backend communication is handled through Tauri Inter-Process Communication. The system consists of three main stages: enrollment, analysis, and export. During enrollment, users register identities through guided multi-pose webcam capture. During analysis, PRIVA detects, recognizes, and tracks faces from an uploaded video. During export, the stored analysis result is used to apply selective blurring. The system follows a privacy-first policy in which faces are blurred by default and kept visible only when the selected enrolled identity is confirmed across tracked frames. During analysis, the input video is conditionally converted into a 720p analysis proxy, while YOLOv8n-Face-960 still uses a fixed 960x960 letterbox input. Frames are sampled using a detector sample rate of 3 and a maximum effective FPS of 30.0. The FPS downsampling factor and effective frame step are calculated using Equations (4) and (5).

$$\text{fps\_downsample\_factor} = \text{ceil}\left(\frac{\text{proxy\_fps}}{30.0}\right) \qquad (4)$$

$$\text{effective\_frame\_step} = \text{sample\_rate} \times \text{fps\_downsample\_factor} \qquad (5)$$

With this configuration, 30 FPS, 60 FPS, and 120 FPS videos are analyzed every 3, 6, and 12 frames, respectively. The sampled workload is processed by up to three worker threads, each with local YOLOv8n-Face-960 and MobileFaceNet ONNX Runtime sessions. An ordered result buffer restores chronological frame order before detections are passed to a single global Deep SORT tracker. At the track level, an identity requires at least 3 positive matching samples for confirmation and at least 5 positive samples with a positive ratio of 0.60 for identity locking. Unconfirmed, unknown, conflicting, or not-yet-locked faces remain blurred by default.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

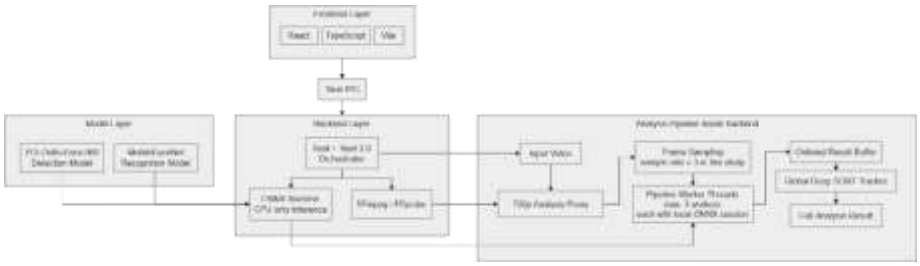


Fig. 3 System design and analysis architecture of PRIVA using YOLOv8n-Face-960, MobileFaceNet, and a global Deep SORT tracker.

System Testing

System-level testing was conducted using five real videos representing vlog-like conditions with variations in face quantity, pose, lighting, movement, occlusion, crowd density, and recording environment. Each video contained at least one enrolled identity and one or more non-enrolled identities. Although PRIVA is designed for Windows deployment, development and testing in this study were conducted in a Linux environment. The evaluation was performed on YOLO analysis frames extracted after temporal sampling and FPS adjustment because these frames represent the actual input processed by YOLOv8n-Face-960. Each analysis frame was manually annotated to identify visible face instances and enrolled or non-enrolled identity status. All videos were processed using the same configuration, consisting of a YOLO confidence threshold of 0.35, NMS threshold of 0.4, MobileFaceNet cosine similarity threshold of 0.40, detector sample rate of 3, maximum effective FPS of 30.0, 720p analysis proxy, and CPU-based ONNX Runtime inference. The first evaluation measured YOLO detection coverage using manually annotated ground truth boxes. A detection was considered correct when the predicted bounding box matched a ground truth face with an IoU value of at least 0.5. Detection performance was calculated using precision, recall, and F1-score.

$$\text{Detection Precision} = \frac{\text{Detection TP}}{\text{Detection TP} + \text{Detection FP}} \times 100\% \quad (6)$$

$$\text{Detection Recall} = \frac{\text{Detection TP}}{\text{Detection TP} + \text{Detection FN}} \times 100\% \quad (7)$$

$$Detection\ F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

The second evaluation measures selective face blurring performance only on correctly detected face instances. In this evaluation, the positive class is a non-enrolled face that should be blurred. TP refers to a non-enrolled face correctly blurred, FP refers to an enrolled face incorrectly blurred, TN refers to an enrolled face correctly kept visible, and FN refers to a non-enrolled face incorrectly kept visible. System performance is calculated using precision, recall, F1-score, and accuracy.

$$\text{System Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (9)$$

$$\text{System Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (10)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

$$\text{System Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (12)$$

In addition to the final PRIVA evaluation, this study also conducts component-level baseline comparisons. MTCNN and YOLOv8n-Face-960 are compared using the same test videos, sampled frames, ground truth boxes, IoU threshold, and detection metrics. FaceNet and MobileFaceNet are compared based on model size and CPU inference time using ONNX Runtime. Processing performance is measured using analysis time, export time, analysis FPS, export FPS, and processing ratio. Finally, blackbox functional testing is conducted to verify the main user-facing workflow from face enrollment to selective blur export.

*name of corresponding author

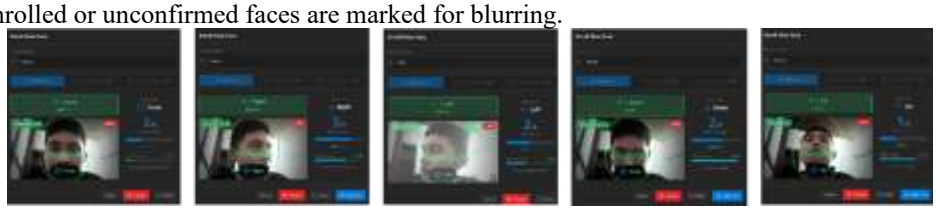
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

This section presents the implementation and evaluation results of PRIVA, including the application interface, test video dataset, YOLO detection coverage, selective face blurring performance, processing performance, and blackbox functional testing.

Implementation Result of PRIVA Application

PRIVA was successfully implemented as a local desktop-based selective face blurring application. The implemented workflow includes face enrollment, video import, automated analysis, review mode, and selective blur export. The application is designed for Windows deployment and was developed and tested in a Linux environment. All AI inference processes run locally through ONNX Runtime without requiring an external AI server. The face enrollment interface allows users to register an identity through guided multi-pose webcam capture. The system captures frontal, right, left, downward, and upward face poses, as shown in Fig. 4(a) to Fig. 4(e). During enrollment, the interface displays the target pose, detected pose, sample count, and enrollment progress. After enrollment, users can import a video, run automated analysis, review detected face tracks and blur decisions, and export the final selectively blurred video. Enrolled faces confirmed by the system are kept visible, while non-enrolled or unconfirmed faces are marked for blurring.



(a) (b) (c) (d) (e)

Fig. 4 The implemented PRIVA multi-pose face enrollment interface shows guided pose capture for (a) frontal, (b) right, (c) left, (d) downward, and (e) upward orientations.

Test Video Dataset Description

The system was evaluated using five real videos representing vlog-like recording conditions. Each video contains at least one enrolled identity and one or more non-enrolled identities, allowing both visible and blurred face decision paths to be evaluated. The dataset includes variations in face quantity, pose, distance, movement, lighting, occlusion, resolution, and recording environment, including low-light, low-resolution, crowded, and multiple-enrolled-identity scenarios. The evaluation was conducted on YOLO analysis frames extracted after temporal sampling and FPS adjustment because these frames represent the actual input processed by YOLOv8n-Face-960. Manual ground truth annotation was performed on these frames. Faces were annotated when they were sufficiently visible and their bounding boxes could be determined consistently. Very distant or indistinct background persons were not counted as unique identities unless their faces were distinguishable in the analysis frames. The test video characteristics are summarized in Table 2.

Table 2. Test video dataset characteristics.

ID	Duration	Resolution	FPS	Total Frames	YOLO Analysis Frames	Enrolled Identities	Non-Enrolled Identities	Max Faces per Frame	Ground Truth Face Instances	Condition
V1	0:23	1440x2560	30	690	232	1	5	3	323	Indoor
V2	0:23	1080x1920	60	1380	238	1	3	2	234	Indoor
V3	0:22	1080x1920	30	660	218	1	6	13	1150	Outdoor, crowded bazaar
V4	0:22	476x848	18.23	404	135	1	4	3	168	Low-light, low-resolution
V5	0:27	1920x1080	30	789	263	2	22	16	3941	Very crowded, multiple

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

										enrolled identities

Overall, the dataset consists of five videos with approximately 117 seconds of duration, 3,923 original frames, 1,086 YOLO analysis frames, and 5,816 manually annotated ground truth face instances.

YOLO Detection Coverage Result

The YOLO detection coverage evaluation was conducted to measure the ability of YOLOv8n-Face-960 to detect visible faces in the extracted analysis frames before selective blur decisions were evaluated. Each analysis frame was manually annotated, and a detection was considered correct when the predicted bounding box matched a ground truth face with an IoU value of at least 0.5. Detection TP refers to a correctly detected visible face, Detection FP refers to a predicted face box where no actual face exists, and Detection FN refers to a visible face missed by YOLOv8n-Face-960. The detection coverage result is shown in Table 3.

Table 3. YOLOv8n-Face-960 detection coverage result.

ID	YOLO Analysis Frames	Ground Truth Face Instances	Detection TP	Detection FP	Detection FN	Detection Precision	Detection Recall	Detection F1-Score
V1	232	323	306	14	17	95.63%	94.74%	95.18%
V2	238	234	226	7	8	97.00%	96.58%	96.79%
V3	218	1150	1031	234	119	81.50%	89.65%	85.38%
V4	135	168	122	7	46	94.57%	72.62%	82.15%
V5	263	3941	3510	10	431	99.72%	89.06%	94.09%

Across all five videos, YOLOv8n-Face-960 produced 5,195 true positives, 272 false positives, and 621 false negatives from 5,816 manually annotated ground truth face instances. Based on these aggregated values, the detector achieved an overall precision of 95.02%, recall of 89.32%, and F1-score of 92.09%. The detector performed strongly in V1 and V2 under indoor conditions, while V4 produced the lowest recall because low-light and low-resolution conditions reduced detection coverage. These results show that YOLOv8n-Face-960 provides strong detection coverage for the PRIVA pipeline, although missed detections still occur in low-quality, distant, crowded, or partially occluded face regions.

Baseline Component Comparison Result

A baseline component comparison was conducted to evaluate whether the final PRIVA components are more suitable than the earlier prototype components. The detection comparison used the same test videos, ground truth annotations, IoU threshold of 0.5, and detection metrics to compare MTCNN and YOLOv8n-Face-960. The recognition comparison evaluated FaceNet and MobileFaceNet based on model size and CPU inference time using ONNX Runtime. For detection, F1-score was used as the comparison metric because it summarizes precision and recall. As shown in Fig. 5, YOLOv8n-Face-960 achieved higher F1-scores than MTCNN across all five test videos.

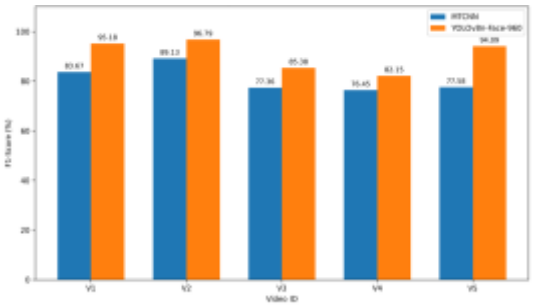


Fig. 5 Detection F1-score comparison between MTCNN and YOLOv8n-Face-960 across the five test videos.

Based on the arithmetic mean of the five videos, YOLOv8n-Face-960 achieved a mean F1-score of 90.72%, while MTCNN achieved 80.84%. This shows an improvement of 9.88 percentage points, which is important because undetected non-enrolled faces cannot be processed by the recognition, tracking, and blur decision stages. The recognition model benchmark was conducted using ONNX Runtime CPUExecutionProvider with 50 warm-

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

up iterations and 1,000 repeated inference iterations on the same benchmark input set. The result is shown in Table 4.

Table 4. Recognition model efficiency comparison between MobileFaceNet and FaceNet.

Model	Model Size	Input Size	Embedding Dimension	Mean Inference Time	Median Inference Time	P95 Inference Time
MobileFaceNet	12.99 MB	112x112	512	4.3499 ms	4.4429 ms	5.2811 ms
FaceNet	89.61 MB	160x160	512	10.0819 ms	11.0141 ms	11.7547 ms

Table 4 shows that MobileFaceNet is more suitable for local CPU-based deployment than FaceNet. The FaceNet ONNX model is approximately 6.90 times larger and 2.48 times slower than MobileFaceNet based on median inference time. These results support the selection of YOLOv8n-Face-960 and MobileFaceNet as the final detection and recognition components in PRIVA.

Selective Face Blurring Performance

The selective face blurring performance evaluation was conducted to measure whether PRIVA correctly applied blur decisions to correctly detected face instances. This evaluation only considers YOLO detections that matched manually annotated ground truth faces. Therefore, missed faces and non-face detections are not counted again in this subsection because they have already been reported as Detection FN and Detection FP in the detection coverage evaluation. In this evaluation, the positive class refers to a non-enrolled face that should be blurred. TP indicates a non-enrolled face correctly blurred, FP indicates an enrolled face incorrectly blurred, TN indicates an enrolled face correctly kept visible, and FN indicates a non-enrolled face incorrectly kept visible. The result is shown in Table 5.

Table 5. Selective face blurring performance result.

Video ID	Correctly Detected Face Instances	TP	FP	TN	FN	System Precision	System Recall	F1-Score	System Accuracy
V1	306	143	0	162	1	100%	99.31%	99.65%	99.67%
V2	226	200	1	25	0	99.50%	100%	99.75%	99.56%
V3	1031	815	0	211	5	100%	99.39%	99.69%	99.52%
V4	122	37	4	77	4	90.24%	90.24%	90.24%	93.44%
V5	3510	2989	18	458	45	99.40%	98.52%	98.96%	98.21%

Across all five videos, PRIVA evaluated 5,195 correctly detected face instances, consisting of 4,184 TP, 23 FP, 933 TN, and 55 FN. Based on these aggregated values, PRIVA achieved an overall system precision of 99.45%, recall of 98.70%, F1-score of 99.08%, and accuracy of 98.50%. The system performed strongly in V1, V2, and V3, while V4 produced the lowest result because low-light and low-resolution conditions affected identity matching and blur decisions. In V5, PRIVA still achieved 98.96% F1-score in a very crowded scenario with multiple enrolled identities. These results show that PRIVA can perform identity-selective face blurring reliably on correctly detected faces, although incorrect decisions may still occur in low-quality or highly crowded scenes.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16273
e-ISSN : 2541-2019
p-ISSN : 2541-044X

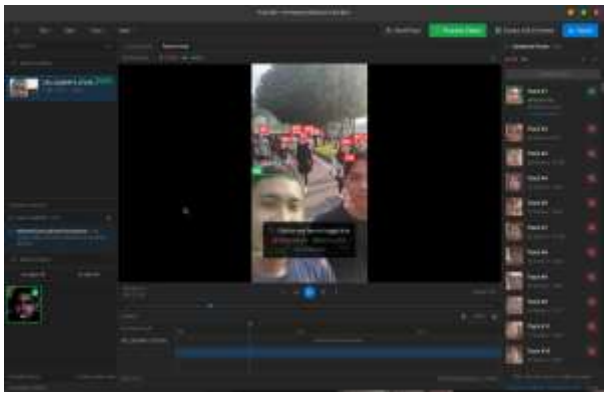


Fig. 6 Example of PRIVA selective face blurring output, showing enrolled faces kept visible while non-enrolled faces are blurred.

Processing Performance Result

The processing performance evaluation was conducted to measure the computational efficiency of PRIVA during video analysis and selective blur export. The test was performed in the Linux development and testing environment using an AMD Ryzen 5 6600H CPU and 14.86 GiB of RAM. Although the device includes an integrated AMD Radeon 680M GPU, all AI inference processes were executed on the CPU through ONNX Runtime. During export, all videos used the pixelate blur type with 100% blur strength. For each video, the recorded attributes include analysis time, export time, analysis FPS, export FPS, and processing ratio. Analysis time refers to detection, recognition, tracking, and blur decision generation, while export time refers to generating the final selectively blurred video. The result is shown in Table 6.

Table 6. PRIVA processing performance result.

ID	Duration	Total Frames	YOLO Analysis Frames	Analysis Time	Export Time	Analysis FPS	Export FPS	Processing Ratio
V1	0:23	690	232	45s	15s	5.16	46.00	2.61x
V2	0:23	1380	238	48s	17s	4.96	81.18	2.83x
V3	0:22	660	218	46s	8s	4.74	82.50	2.45x
V4	0:22	404	135	30s	5s	4.50	80.80	1.58x
V5	0:27	789	263	56s	11s	4.70	71.73	2.47x

The result in Table 6 shows that the analysis stage required 30 to 56 seconds, with analysis FPS ranging from 4.50 to 5.16 frames per second. This indicates that CPU-only inference requires more processing time than the original video duration because the system performs detection, recognition, tracking, and blur decision generation. The export stage was faster, with export FPS ranging from 46.00 to 82.50 frames per second, because it mainly applies stored blur decisions to the original video frames. Across all videos, PRIVA achieved an average analysis speed of 4.83 FPS, average export speed of 70.05 FPS, and an average processing ratio of approximately 2.40x the original video duration.

Blackbox Functional Testing Result

Blackbox functional testing was conducted to verify whether the main user-facing features of PRIVA operate correctly based on observable input and output. The tested features include face enrollment, face library storage, video import, automated analysis, review mode, selective blur decision display, and video export. A feature was considered passed when the observed behavior matched the expected function. The result is shown in Table 7.

Table 7. Blackbox functional testing result.

Tested Feature	Expected Function	Status
Face enrollment initialization	Opens the webcam enrollment interface and displays pose instructions.	Passed

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Multi-pose face capture	Captures 20 face samples across the required pose directions.	Passed
Face library storage	Saves the enrolled identity and displays it in the face library.	Passed
Video import	Loads the selected video into the editor workspace with metadata.	Passed
Automated video analysis	Processes the video and returns detected face tracks with blur decisions.	Passed
Review mode display	Displays bounding boxes, track information, and blur decision overlays.	Passed
Selective blur decision	Marks enrolled faces as visible and non-enrolled or unconfirmed faces for blur.	Passed
Selective blur export	Generates an output video with selective face blurring applied.	Passed

The blackbox testing result shows that all tested features passed the expected behavior criteria, confirming that the main PRIVA workflow can be executed from face enrollment to selective blur export.

DISCUSSIONS

The evaluation results show that PRIVA can perform identity-selective face blurring under the tested vlog-like video conditions. YOLOv8n-Face-960 achieved an overall detection precision of 95.02%, recall of 89.32%, and F1-score of 92.09%, indicating strong detection coverage across five test videos. However, missed detections still occurred in low-light, low-resolution, crowded, distant, or partially occluded face regions. This limitation is important because undetected non-enrolled faces cannot be processed by the recognition and blur decision stages. The baseline component comparison supports the final component selection. YOLOv8n-Face-960 achieved a higher mean detection F1-score than MTCNN, while MobileFaceNet was smaller and faster than FaceNet for CPU-based local inference. For correctly detected face instances, PRIVA achieved 99.45% precision, 98.70% recall, 99.08% F1-score, and 98.50% accuracy, showing that the combination of MobileFaceNet, Deep SORT, and blur decision logic can reliably distinguish enrolled and non-enrolled faces after detection. The system also achieved an average analysis speed of 4.83 FPS, export speed of 70.05 FPS, and processing ratio of approximately 2.40x the original video duration. Blackbox testing confirmed that the main application workflow operated successfully. Overall, PRIVA is functionally feasible as a local desktop privacy editing tool, although future work should improve detection robustness, expand the dataset, and test more varied identities and recording conditions.

CONCLUSION

This study successfully developed PRIVA as a local desktop-based selective face blurring application using YOLOv8n-Face-960, MobileFaceNet, and Deep SORT. The system provides guided multi-pose enrollment, video analysis, review mode, and selective blur export, allowing enrolled faces to remain visible while non-enrolled or unconfirmed faces are blurred. Evaluation on five real videos showed that YOLOv8n-Face-960 achieved 95.02% precision, 89.32% recall, and 92.09% F1-score for face detection. The baseline comparison showed that YOLOv8n-Face-960 outperformed MTCNN in mean F1-score, while MobileFaceNet was smaller and faster than FaceNet for CPU inference. For correctly detected faces, PRIVA achieved 99.45% precision, 98.70% recall, 99.08% F1-score, and 98.50% accuracy in selective blur decisions. The CPU-based pipeline achieved 4.83 FPS analysis speed, 70.05 FPS export speed, and approximately 2.40x processing ratio. These results indicate that PRIVA can support practical local video privacy editing without external AI servers, although future work is needed to improve robustness in low-light, low-resolution, crowded, distant, and partially occluded face conditions.

REFERENCES

Chen, S., Liu, Y., Gao, X., & Han, Z. (2018). MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices. *Biometric Recognition (CCBR 2018), Lecture Notes in Computer Science*, *10996*, 428–438. https://doi.org/10.1007/978-3-319-97909-0_46

European Parliament, & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). In *Official Journal of the European Union: L 119* (pp. 1–88). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

Google LLC. (2024). *Privacy guidelines: Protecting your identity*. YouTube Help Center. https://support.google.com/youtube/answer/2801895

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

Kemp, S. (2025). *Digital 2025: Indonesia*. DataReportal. https://datareportal.com/reports/digital-2025-indonesia

*name of corresponding author

Khan, W., Topham, L., Khayam, U., Ortega-Martorell, S., Heather, P., Ansell, D., Al-Jumeily, D., & Hussain, A. (2025). Person de-identification: A comprehensive review of methods, datasets, applications, and ethical aspects along with new dimensions. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *7*(3), 293–312. https://doi.org/10.1109/TBIOM.2024.3485990

Kosasih W., P., & Engel, M. M. (2026). Performance Comparison of Tauri and Electron Frameworks in Multiplatform Desktop Application Development. *Bit-Tech*, *8*(3), 3875–3882. https://doi.org/10.32877/bt.v8i3.3733

Laishram, L., Lee, J. T., & Jung, S. K. (2024). Face De-Identification Using Face Caricature. *IEEE Access*, *12*, 19344–19354. https://doi.org/10.1109/ACCESS.2024.3356550

Laishram, L., Shaheryar, M., Lee, J. T., & Jung, S. K. (2024). Toward a Privacy-Preserving Face Recognition System: A Survey of Leakages and Solutions. *ACM Computing Surveys*. https://doi.org/10.1145/3673224

Meta. (2024). *Community Standards*. https://transparency.meta.com/policies/community-standards/privacy-violations/

Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, *54*(1), 1–41. https://doi.org/10.1145/3425780

Plaud, R., & Lisani, J.-L. (2024). Two Deep Learning Solutions for Automatic Blurring of Faces in Videos. *The Second Tiny Papers Track at ICLR 2024*. https://arxiv.org/abs/2409.14828

Putra, B. K., Putrada, A. G., & Oktaviani, I. D. (2025). *YOLOv8 and FaceNet Algorithm for Real-Time Face Recognition Attendance System*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. http://arxiv.org/abs/1506.02640

Republik Indonesia. (2022). *Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*. Lembaran Negara Republik Indonesia Tahun 2022 Nomor 196. https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

TikTok. (2024). *Privacy Policy*. https://www.tiktok.com/legal/page/global/privacy-policy/en

Wang, M., Li, S., & Zhang Xinpeng and Feng, G. (2025). Facial Privacy in the Digital Era: A Comprehensive Survey on Methods, Evaluation, and Future Directions. *Computer Science Review*, *58*, 100785. https://doi.org/10.1016/j.cosrev.2025.100785

We Are Social, & Hootsuite. (2023). *Digital 2023: Indonesia*. We Are Social. https://wearesocial.com/id/blog/2023/01/digital-2023/

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple Online and Realtime Tracking with a Deep Association Metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962

Xu, D., Chen, T., Pearce, J., Mohammadi, Z., & Pearce, P. L. (2021). Reaching audiences through travel vlogs: The perspective of involvement. *Tourism Management*, *86*. https://doi.org/10.1016/j.tourman.2021.104326

*name of corresponding author