Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Comparative Evaluation of YOLOv5 and YOLOv8 Models in Detecting Smoking Behavior

Muhammad Andaru Megaarta[1]*

[1]*Politeknik Negeri Sriwijaya*
andarumega05@gmail.com[1*]

Abstract

Smoking behavior in public spaces has become a serious concern in public health efforts, as it poses health risks not only to active smokers but also to passive smokers. This study presents a comparative evaluation of two state-of-the-art object detection models, YOLOv5 and YOLOv8, for the automatic detection of smoking behavior. The models were trained on a labeled image dataset containing cigarettes, faces, and smoking activities. Evaluation metrics used in this study include precision, recall, F1-score, and mean Average Precision (mAP). The experimental results show that both models achieved strong detection performance, with precision, recall, and F1-scores above 0.95. YOLOv5 obtained slightly higher precision (0.98064), recall (0.96388), and F1-score (0.97), while YOLOv8 achieved a marginally higher mAP (0.97782), indicating better generalization across varying IoU thresholds. YOLOv8 also showed improved classification performance in detecting faces (0.69) and smoking behavior (0.54), benefiting from its anchor-free architecture and advanced loss functions. These findings demonstrate that while both models are highly effective, YOLOv8 offers greater robustness and accuracy for real-time smoking detection in complex public environments, supporting efforts to minimize cigarette exposure and improve public health awareness.

Keywords: *Smoking Behavior Detection, YOLOv5, YOLOv8, Object Detection, Deep learning.*

1. Introduction

Smoking behavior in public spaces has become a pressing public health concern that demands serious attention. Exposure to cigarette smoke poses health risks not only to active smokers but also to those around them, including passive smokers. As such, the ability to automatically detect and monitor smoking behavior is essential for supporting tobacco control policies and raising public awareness about the dangers of smoking. [1]. Traditionally, smoking detection and monitoring have relied on manual observation and self-reported surveys, which are often time-consuming, subjective, and prone to reporting bias. These conventional methods also lack the capability for continuous and large-scale monitoring in public spaces, limiting their effectiveness in enforcing smoking regulations. [2].

Advancements in artificial intelligence, particularly in computer vision and object detection, have opened up new opportunities for developing real-time and accurate smoking detection systems. [3]. Among the most widely used methods is the YOLO (You Only Look Once) algorithm, renowned for its speed and accuracy in identifying objects in images and video streams [4]. Currently, YOLOv5 and YOLOv8 are two of the most popular object detection models used across a wide range of applications. [5]. YOLOv8 introduces several architectural innovations, including an anchor-free detection mechanism and more efficient convolutional modules, which have the potential to enhance both detection accuracy and computational efficiency compared to YOLOv5. However, a comprehensive performance comparison of these two models, specifically in the context of smoking behavior detection, remains limited.

One recent study by Paramita et al. (2024). [6] addressed the issue of tobacco product imagery on social media, which can influence youth by glamorizing smoking, by comparing the performance of YOLOv5 and YOLOv8 models for cigarette detection and censorship. Using a dataset of 2,188 Twitter images, they trained and evaluated both models with metrics such as mean Average Precision (mAP) and F1-Score, finding that YOLOv8 achieved a slightly higher mAP (0.933) than YOLOv5 (0.919), as well as improved precision and recall, highlighting YOLOv8's architectural advantages and anchor-free detection mechanism. Their research demonstrated that both models had robust learning without overfitting, and they concluded that automated censorship using deep learning, particularly with YOLOv8, could significantly reduce youth exposure to tobacco-related content on social media, thus contributing to digital public health efforts.

This study aims to conduct a comparative evaluation of YOLOv5 and YOLOv8 for detecting smoking behavior using a relevant image dataset. The evaluation is based on standard performance metrics, including mean Average Precision (mAP), precision, recall, and F1-score, to assess the accuracy and reliability of both models. The findings of this research are expected to offer valuable insights into the strengths and limitations of each model, providing a foundation for the development of more effective and efficient public-space smoking behavior monitoring systems.

2. Research Methodology

The stages in this research method are designed to compare the performance of the YOLOv5 and YOLOv8 models in detecting smoking behavior. The process begins with dataset collection, followed by data preprocessing, model training, testing, validation, and finally, evaluation to select the best-performing model.

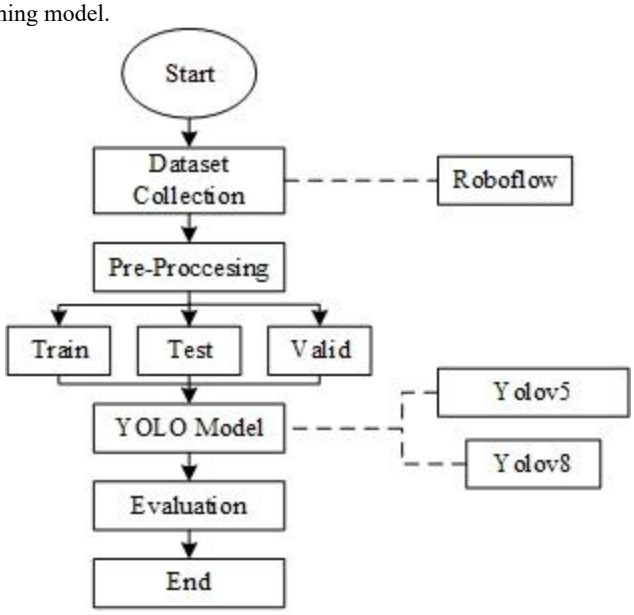

Fig. 1: Research WorkFlow

2.1. Data Colllection

This study involves the collection of an image dataset to be used for training the smoking behavior detection model. The dataset can be sourced from various platforms, one of which is Roboflow, which provides the "Smoking Dataset" [7]. This dataset contains 2,337 labeled images, categorized into three object classes: face, cigarette, and smoker. These categories are essential for training the object detection model, as they enable the system to recognize relevant elements within the context of smoking behavior. The labeling process helps the model to understand and distinguish between individual faces, the presence of cigarettes, and the identification of smokers in the analyzed images.

2.2. Data Preprocessing

After the dataset has been collected, the next step is preprocessing. This process begins by resizing the images from their original dimensions to a standard size of 640 × 640 pixels [3]. Adjusting the image size is crucial to ensure data consistency, facilitate the model training process, and reduce the variation in image sizes that could affect model performance [8]. Additionally, resizing helps optimize memory usage and processing speed, as the model will handle images of more uniform dimensions. Once the images are resized, further preprocessing steps are carried out to prepare the data for training.

2.3. YOLO Architecture

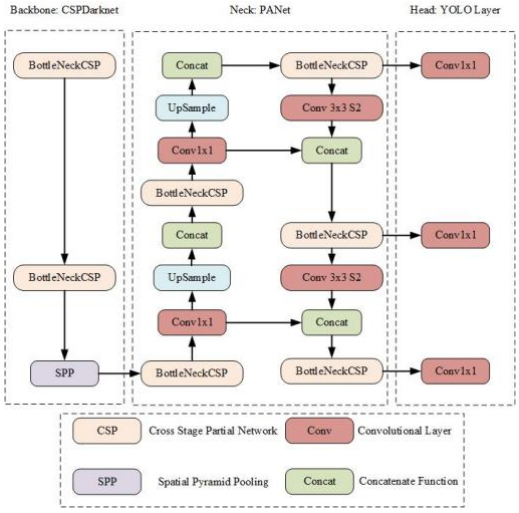

Fig. 2: YOLOv5 Architecture

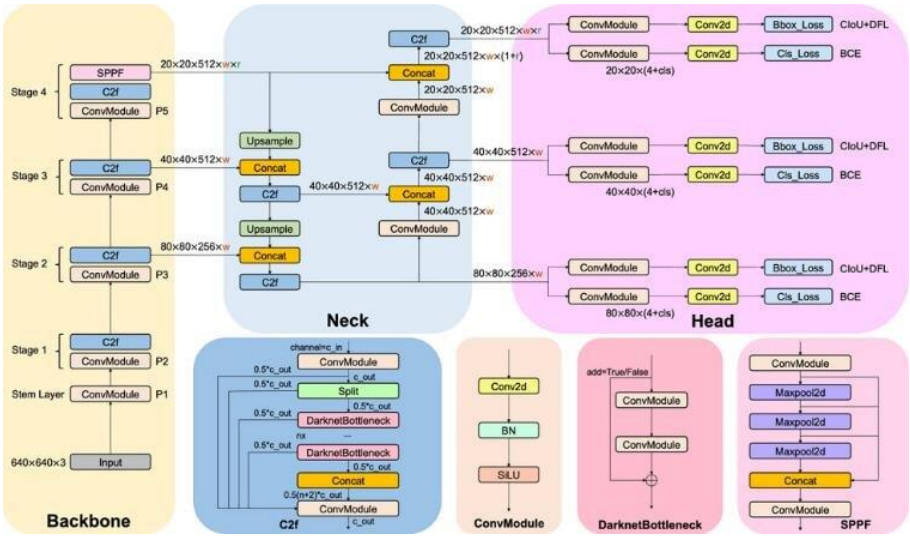

Fig. 3: YOLOv8 Architecture

YOLOv5 and YOLOv8 are state of the art object detection models that share a similar overall structure composed of three main parts: the backbone, the neck, and the head. However, YOLOv8 introduces several architectural innovations that improve accuracy and efficiency compared to YOLOv5.

YOLOv5 uses the CSPDarknet architecture as its backbone, which incorporates Cross Stage Partial connections to enhance feature extraction while maintaining computational efficiency [9]. YOLOv8 builds upon this by enhancing the CSPDarknet backbone with the introduction of the C2f module, which replaces the C3 module used in YOLOv5. The C2f module integrates all bottleneck outputs more effectively and increases the convolution kernel size from 1×1 to 3×3 in the first convolution layer, resulting in better feature representation and improved accuracy. [10].

Both YOLOv5 and YOLOv8 utilize a neck component to aggregate multi-scale features, critical for detecting objects of varying sizes [11]. YOLOv5 employs PANet (Path Aggregation Network) to enhance information flow between different feature levels [12]. YOLOv8 improves upon this by using an optimized PANet variant with enhanced cross-layer connections, which allows for more efficient feature fusion and better spatial information preservation, contributing to higher detection precision. [13].

The detection head is responsible for predicting bounding boxes, objectness scores, and class probabilities. [14]. YOLOv5 uses an anchor-based detection mechanism where bounding boxes are predicted relative to predefined anchor boxes. [3]. In contrast, YOLOv8 adopts an anchor-free detection approach, directly predicting the center points and sizes of bounding boxes without relying on anchor boxes. [15]. This simplification reduces complexity, speeds up training, and often improves detection accuracy, especially for objects with varying shapes and sizes.

YOLOv8 also introduces a split head design, separating classification and bounding box regression tasks, unlike YOLOv5's single-head approach. This separation allows for more specialized learning and better performance. [16]. Additionally, YOLOv8 uses advanced loss functions such as Tangent-Aided Loss (TAL) and Dynamic Focal Loss (DFL), which further improve training stability and accuracy compared to YOLOv5's focal and IoU losses. [17].

Overall, YOLOv8 achieves higher mean Average Precision (mAP) scores than YOLOv5 across various model sizes while maintaining competitive inference speeds on GPUs. Although YOLOv8 may have slightly higher CPU inference times in some cases, its improved accuracy and efficiency on optimized hardware make it preferable for real-time applications

2.4. Model Evaluation

Model evaluation is a crucial stage in assessing the performance of a machine learning model, especially in classification and object detection tasks. Several metrics are commonly used to provide a comprehensive evaluation, including precision, recall, mean Average Precision (mAP), and F1 score. [18].

$$Precision = \frac{TP}{TP+ FP} \tag{1}$$

$$Recall = \frac{TP}{TP+ FN} \tag{2}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

$$mAP = \frac{1}{N}\sum_{i=0}^{N} APi \tag{4}$$

3. Result and Discussion

3.1. Training Result

The training results are presented and analyzed in this chapter, supported by graphical representations to illustrate the model's learning progress and performance over time. The graphs provide a clear visualization of key metrics such as training loss, validation loss, enabling a comprehensive understanding of how the model improves through each epoch. These visual aids help to identify trends, potential overfitting or underfitting, and the overall effectiveness of the training process. Detailed discussion of the graphs follows to interpret the implications of the observed patterns on the model's capability.

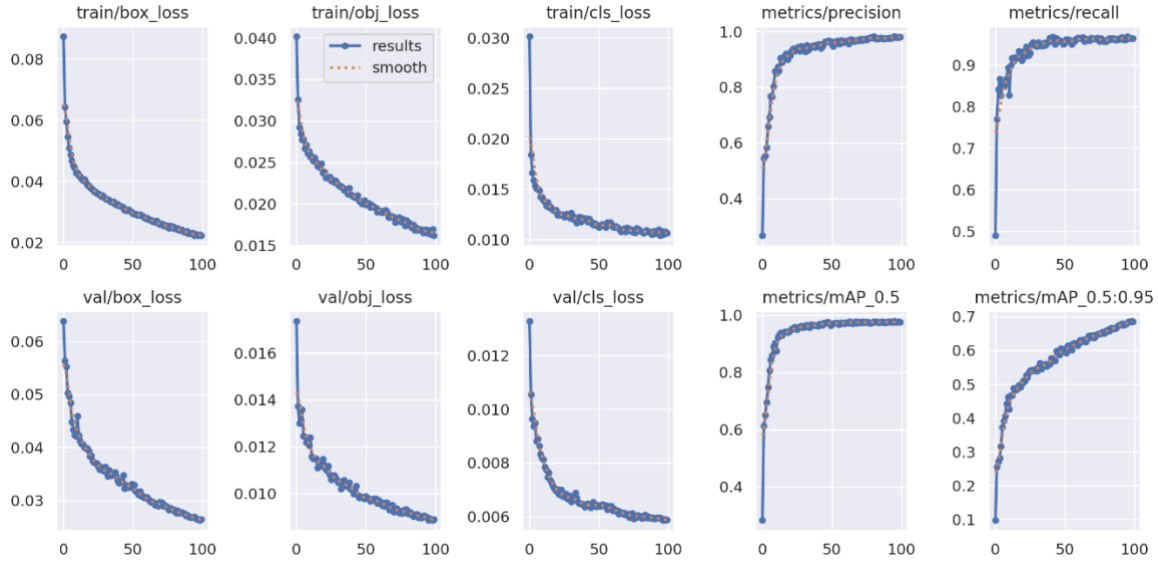

Fig. 4: Loss and Accuracy Graphs of the YOLOv5 Model

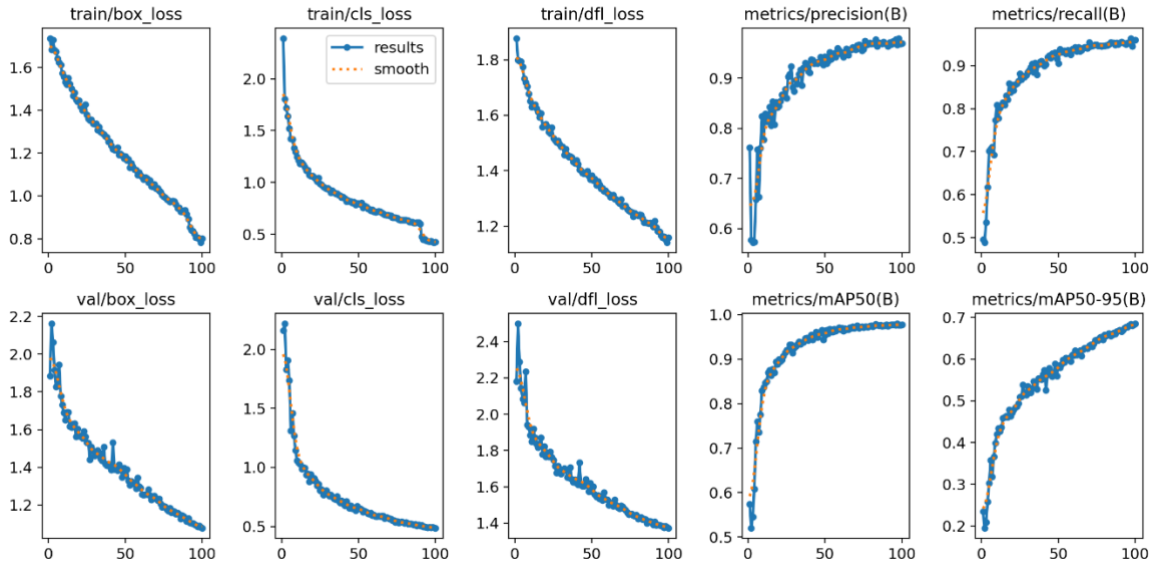

Fig. 5: Loss and Accuracy Graphs of the YOLOv8 Model

The main differences between the training result graphs of YOLOv5 and YOLOv8 stem primarily from the distinct loss functions they employ and the architectural improvements introduced in YOLOv8. YOLOv5 uses an anchor-based detection mechanism, which includes an objectness loss component that measures how confidently the model predicts the presence of an object within predefined anchor boxes. This objectness loss helps the model distinguish between foreground objects and background, but it also adds complexity to the training process. In contrast, YOLOv8 adopts an anchor-free detection approach, eliminating the need for anchor boxes and replacing the objectness loss with a Distribution Focal Loss (DFL). This loss function focuses on improving the precision of bounding box regression by modeling the distribution of bounding box offsets more effectively, leading to more accurate localization.

The shift from YOLOv5's anchor-based detection with objectness loss to YOLOv8's anchor-free approach using Distribution Focal Loss (DFL) leads to distinct training behaviors; YOLOv5's objectness loss steadily decreases as the model learns to separate objects from the background, whereas YOLOv8's DFL loss may start higher but converges as bounding box predictions improve, reflecting its architectural innovation to simplify the detection pipeline and enhance localization accuracy. Additionally, YOLOv8 introduces a split head design that separates classification and bounding box regression tasks, enabling more specialized learning and better performance. Combined with advanced loss functions like DFL and Tangent-Aided Loss (TAL), these enhancements improve training stability and accuracy beyond YOLOv5's focal and IoU losses. Empirical results show that YOLOv8 generally achieves higher mean Average Precision (mAP) across various datasets, indicating superior detection accuracy, while maintaining competitive inference speed with fewer or comparable parameters and FLOPs. This balance is evident in training graphs where YOLOv8's precision, recall, and mAP stabilize at higher values, demonstrating its enhanced capability in object detection. Overall, the transition to anchor-free detection and sophisticated loss functions enables YOLOv8 to outperform YOLOv5 in bounding box accuracy and overall detection effectiveness, as clearly reflected in their respective training loss and metric graphs.

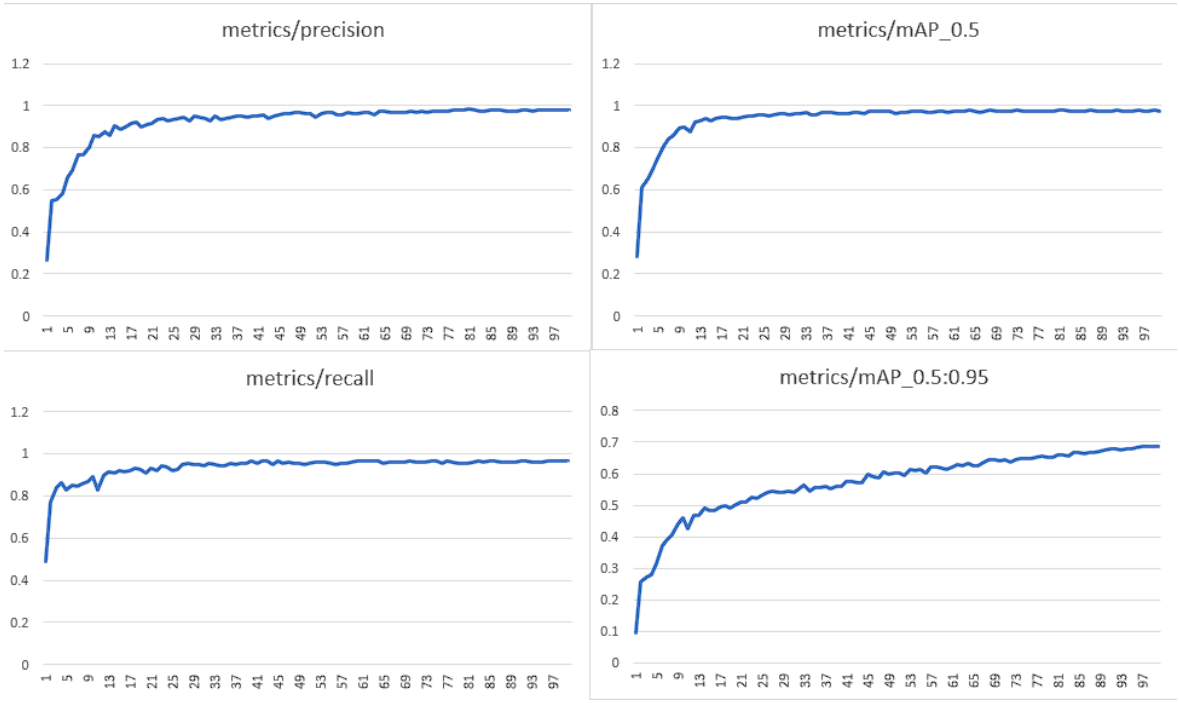

Fig. 6: YOLOv5 Performance Result

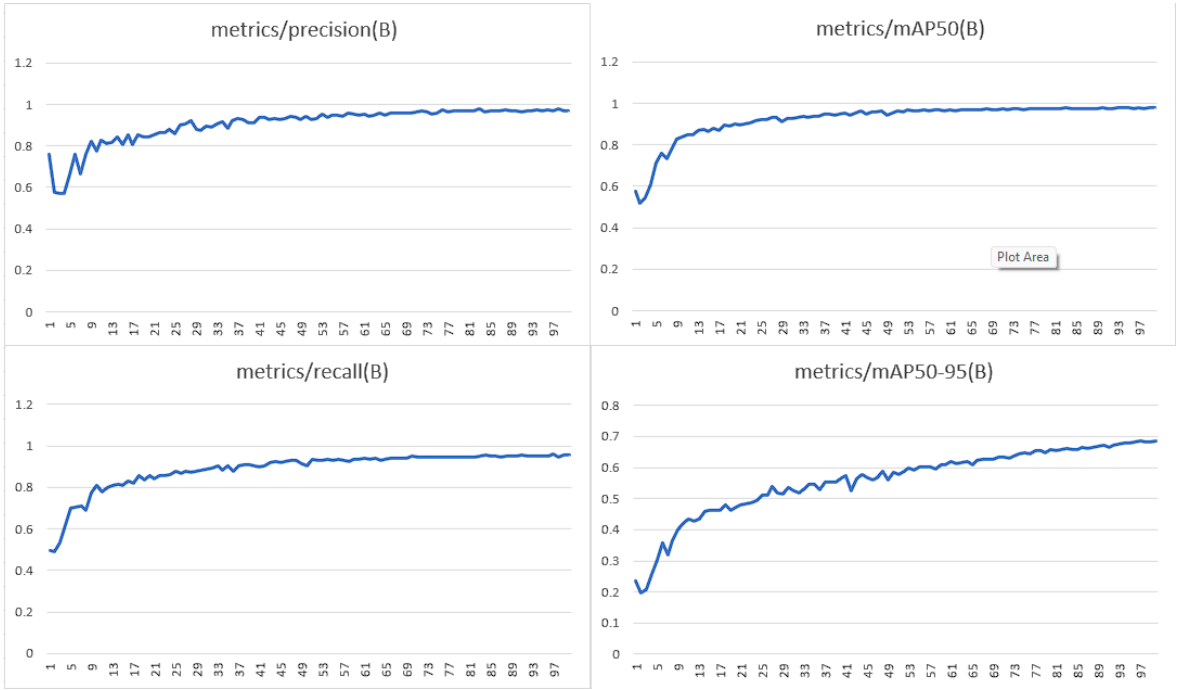

Fig. 7: YOLOv8 Performance Result

The training result graphs for YOLOv5 and YOLOv8 illustrate the performance progression of each model across several key evaluation metrics: precision, recall, mAP 0.5, and mAP 0.5:0.95. From the visual analysis, both models demonstrate strong learning behavior, but there are some notable differences in their performance trends.

In terms of precision, YOLOv5 starts off with a relatively high value, indicating that it was already making fairly accurate predictions early in the training process. The precision curve for YOLOv5 rises quickly and stabilizes near 0.98. On the other hand, YOLOv8 begins with a lower initial precision, around 0.6, but improves rapidly and eventually reaches a similar level of performance, converging close to 0.99. This suggests that while YOLOv5 may converge more quickly, YOLOv8 is capable of catching up and matching its performance in the long run.

Looking at recall, a similar pattern is observed. YOLOv5 again shows a stronger starting point, with recall beginning around 0.85 and increasing steadily. YOLOv8 starts from a lower recall value, roughly 0.6, but improves consistently and ends up with nearly the same recall score as YOLOv5. This demonstrates YOLOv8's ability to generalize well across training epochs, despite needing more time to reach peak performance.

For the mAP 0.5 metric, which measures the average precision at a fixed IoU threshold of 0.5, both models perform very similarly. YOLOv5 and YOLOv8 reach high mAP values above 0.95 early in training and maintain them steadily through the remaining epochs. This indicates that both models are highly capable of detecting smoking-related objects when only a moderate degree of overlap (IoU) is required.

However, the mAP 0.5:0.95, a stricter and more comprehensive metric that accounts for detection quality across a range of IoU thresholds, reveals a slight advantage for YOLOv8. While both models start with low values around 0.1, YOLOv8 shows more stable and consistent improvement, ending with a slightly higher final value (approximately 0.69) compared to YOLOv5 (approximately 0.67). This suggests that YOLOv8 is slightly better at maintaining accuracy across more difficult detection conditions, likely due to its architectural improvements such as anchor-free detection and more efficient feature extraction.

Overall, both models show strong and reliable performance in detecting smoking behavior, with YOLOv5 offering faster convergence and YOLOv8 delivering slightly higher accuracy under complex evaluation settings. Neither model shows signs of overfitting or underfitting, and the training curves indicate that both are well-trained and stable by the final epochs.

3.2. Model Evaluation

Following the completion of model training and selection of the optimal model, the evaluation phase involves testing to generate the confusion matrix. This matrix serves as the basis for calculating important performance metrics including accuracy, precision, recall, and F1-score. The subsequent section details the testing outcomes through the presentation of these confusion matrices.

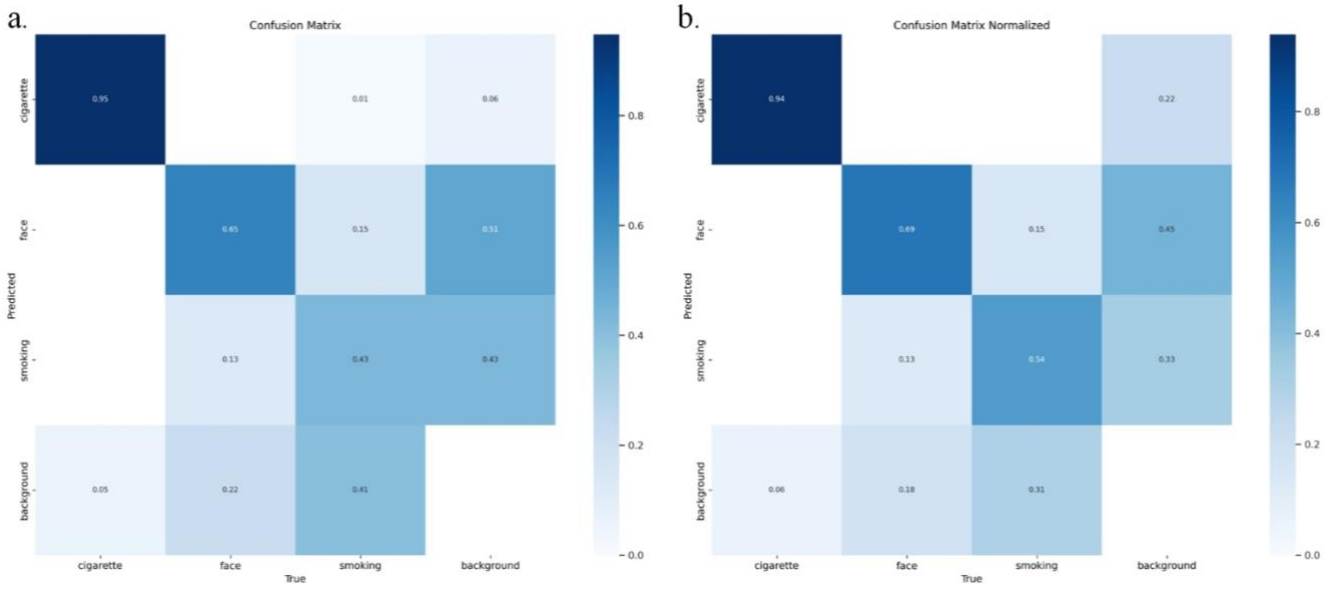

Fig. 8: Confusion Matrix a) YOLOv5, b) YOLOv8

Figure a present the confusion matrix for the YOLOv5 model, while figure b shows the normalized confusion matrix for the YOLOv8 model. Both matrices evaluate the models' abilities to classify four categories: cigarette, face, smoking, and background. In both cases, the highest values are found along the diagonal, indicating correct predictions for each class. For YOLOv5, the model demonstrates strong performance in detecting cigarettes, with a correct classification rate of 0.95. However, there is still some confusion, as a small proportion of cigarette instances are misclassified as background or face. The face class is correctly identified 63% of the time, but is sometimes confused with the smoking and background classes. The smoking class shows a moderate correct classification rate of 0.43, with notable confusion particularly with the background and face classes. The background class is correctly classified 41% of the time, but is frequently misidentified as face or smoking.

In comparison, the YOLOv8 confusion matrix (figure b) indicates an improvement in several categories. The model maintains high accuracy for cigarette detection at 0.94, similar to YOLOv5. Notably, YOLOv8 shows better performance in identifying faces, with a correct classification rate of 0.69, and also achieves a higher recall for the smoking class at 0.54. However, some confusion between background and smoking remains, as indicated by the off-diagonal values. Overall, YOLOv8 demonstrates improved capability in distinguishing between the face and smoking classes compared to YOLOv5, resulting in a more balanced classification across all categories. These results suggest that YOLOv8 offers enhanced performance for multi-class detection in the context of cigarette-related object recognition.

Fig. 9: Model detection result a) YOLOv5 and b) YOLOv8

The visual comparison in the image presents detection results from two different models: YOLOv5 on the left side (labeled a.) and YOLOv8 on the right side (labeled b.). Both models were tasked with identifying objects associated with smoking behavior, particularly cigarettes, faces, and smoking activity.

From the YOLOv5 results (left), the model demonstrates consistent detection of cigarettes across various image contexts. The red bounding boxes clearly highlight smoking-related objects, and in many cases, the model also tags the presence of a face. However, YOLOv5 occasionally shows overlapping boxes and redundant detections, particularly when identifying multiple objects in close proximity, such as a cigarette and face in the same region.

On the right, the YOLOv8 model displays improved localization and labeling performance. The blue and cyan bounding boxes indicate a cleaner, more refined detection outcome. YOLOv8 more accurately isolates individual objects and demonstrates better separation between overlapping instances. Notably, it performs well on both grayscale and low-light images, maintaining label consistency. The detection of "smoking" as a behavior appears more precise in YOLOv8, with less noise and fewer false positives compared to YOLOv5.

Overall, while both models are effective in identifying smoking-related elements, YOLOv8 clearly exhibits superior visual clarity, more accurate object boundaries, and better generalization across diverse image condition, Reinforcing its architectural advantages and anchor-free design.

Table 1: Comparison of model evaluation

Model	Precision	Recall	F1-Score	mAP
YOLOv5	0.98064	0.96388	0.97	0.97503
YOLOv8	0.96875	0.95978	0.96	0.97782

Table 1 presented provides a quantitative comparison between the YOLOv5 and YOLOv8 models based on four key evaluation metrics, Precision, Recall, F1-Score, and mean Average Precision (mAP).

From the results, YOLOv5 demonstrates slightly higher values in precision (0.98064 vs. 0.96875), recall (0.96388 vs. 0.95978), and F1-score (0.97 vs. 0.96), indicating that it performs marginally better in detecting relevant smoking-related objects with fewer false positives and false negatives. These metrics suggest that YOLOv5 is slightly more balanced in recognizing true smoking instances while minimizing errors.

However, YOLOv8 achieves a slightly higher mAP value (0.97782) compared to YOLOv5 (0.97503). Since mAP measures the overall precision and recall across different Intersection over Union (IoU) thresholds, this result suggests that YOLOv8 may offer more consistent detection accuracy across varying object shapes, sizes, and conditions.

Despite the marginal differences, both models perform at a very high level, with metrics close to or above 95%. YOLOv5 leads in traditional detection metrics, while YOLOv8's improved mAP reflects its stronger generalization and robustness—likely due to its anchor-free architecture and more advanced design features. These results affirm that both models are suitable for smoking behavior detection, with YOLOv8 showing promise for broader and more dynamic applications.

4. Conclusion

This study conducted a comparative analysis of YOLOv5 and YOLOv8 models for detecting smoking behavior using an image dataset containing cigarettes, faces, and smoking activities. Both models achieved high accuracy, with precision, recall, F1-score, and mAP values exceeding 95%, indicating their strong capability in object detection. YOLOv5 demonstrated slightly better results in precision (0.98064), recall (0.96388), and F1-score (0.97), suggesting more stable early convergence and lower error rates during training. However, both models showed no signs of overfitting and maintained consistent performance throughout the evaluation.

On the other hand, YOLOv8 achieved a slightly higher mAP (0.97782), supported by its anchor-free detection approach and enhanced architectural features such as the C2f module, split-head design, and advanced loss functions. These improvements led to better bounding box localization, improved detection of faces (0.69 accuracy), and higher recall for smoking classification (0.54), as shown in the confusion matrix and visual results. While both models are reliable and efficient, YOLOv8 proved to be more robust and accurate under complex scenarios, making it more suitable for real-time implementation in public health surveillance systems aimed at reducing smoking exposure.

Acknowledgement

I would like to express my sincere gratitude to all those who have provided valuable support and assistance throughout the completion of this work, especially to Laila Syahrani for her continuous encouragement and unwavering support, as well as to my beloved parents and family, including my mother, father, and all those closest to me, whose unconditional love, prayers, and belief in me have been a constant source of strength.

References

[1] S. A. Putri, “Implementasi Algoritma YOLOv5 untuk Otomatisasi Iklan Layananan Publik tentang Larangan Merokok,” pp. 195–202, 2023.
[2] A. M. Fathoni, E. Zuliarso, and U. S. Semarang, “IMPLEMENTATION OF YOLOv5 METHOD IN THE CIGARETTE DETECTION,” vol. 7, pp. 1449–1454, 2024.
[3] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018, [Online]. Available: http://arxiv.org/abs/1804.02767
[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” 2020, [Online]. Available: http://arxiv.org/abs/2004.10934
[5] I. P. Sary, S. Andromeda, and E. U. Armin, “Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images,” *Ultim. Comput. J. Sist. Komput.*, vol. 15, no. 1, pp. 8–13, 2023, doi: 10.31937/sk.v15i1.3204.
[6] C. Paramita, C. Supriyanto, and K. Rahmyanto Putra, “Comparative Analysis of YOLOv5 and YOLOv8 Cigarette Detection in Social Media Content,” *Sci. J. Informatics*, vol. 11, no. 2, pp. 341–352, 2024, doi: 10.15294/sji.v11i2.2808.
[7] yeolduri, “smoking Dataset,” Mar. 2025, *Roboflow*. [Online]. Available: https://universe.roboflow.com/yeolduri/smoking-hvni4
[8] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
[9] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, “CSPNet: A new backbone that can enhance learning capability of CNN,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2020-June, no. June, pp. 1571–1580, 2020, doi: 10.1109/CVPRW50498.2020.00203.
[10] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, “A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS,” *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.
[11] X. Li, T. Lai, S. Wang, Q. Chen, C. Yang, and R. Chen, “Weighted feature pyramid networks for object detection,” *Proc. - 2019 IEEE Intl Conf Parallel Distrib. Process. with Appl. Big Data Cloud Comput. Sustain. Comput. Commun. Soc. Comput. Networking, ISPA/BDCloud/SustainCom/SocialCom 2019*, pp. 1500–1504, 2019, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217.
[12] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path Aggregation Network for Instance Segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8759–8768, 2018, doi: 10.1109/CVPR.2018.00913.
[13] Z. Guo, W. Zhang, Z. Liang, Y. Shi, and Q. Huang, “Multi-Scale Object Detection Using Feature Fusion Recalibration Network,” *IEEE Access*, vol. 8, pp. 51664–51673, 2020, doi: 10.1109/ACCESS.2020.2980737.
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.
[15] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, “CenterNet: Keypoint triplets for object detection,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-October, pp. 6568–6577, 2019, doi: 10.1109/ICCV.2019.00667.
[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.
[17] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU loss: Faster and better learning for bounding box regression,” *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, no. February, pp. 12993–13000, 2020, doi: 10.1609/aaai.v34i07.6999.
[18] M. M. Sebatubun and C. Haryawan, “Implementasi Algoritma Convolutional Neural Network untuk Klasifikasi Jenis Keris,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 3, pp. 595–602, 2024, doi: 10.25126/jtiik.937260.