Face Detection with Landmark using YOLOv8

Nehith Sai Vemulapalli[1], Paladugula Pruthvi[1], Guggiliam Sai Prabhat[1],
S Abhishek[2], Anjali T [3]
amenu4aie20053@am.students.amrita.edu[1], amenu4aie20054@am.students.amrita.edu[1]
amenu4aie20032@am.students.amrita.edu[1], sabhishek@am.students.amrita.edu[2] anjalit@am.amrita.edu[3]
[1 2 3]Department of Computer Science and Engineering, Amrita School of Computing,
Amrita Vishwa Vidyapeetham, Amritapuri, India

Abstract*—The advent of deep learning combined with computer vision has brought forth unparalleled advancements in facial detection and landmark identification. One pivotal player in this transformation has been the YOLO (You Only Look Once) series, setting new milestones in object detection methodologies. Our research is centered on harnessing the YOLOv8 model to optimize face detection processes. We incorporate the OpenCV library for image processing, enhancing detection fidelity through adjustable parameters like confidence and intersection over union (IoU) thresholds. A standout feature of our methodology is its innate ability to adjust to diverse image proportions. This is achieved by innovatively resizing and padding input images, which not only maintains consistency in detection but also augments accuracy. The proposed technique not only demarcates the face but also pinpoints facial landmarks, thus offering a comprehensive spatial map for each detected face. Preliminary results on benchmark datasets underscore the model's dual advantage of speed and precision. Our approach promises not only improved face detection but also paves the way for its amalgamation into expansive facial recognition and analytic platforms.

***Index Terms*—YOLOv8, Face Detection, Landmark Identification, OpenCV, Image Adaptability**

I. Introduction

In the continuously advancing world of computer vision, the confluence of facial detection and landmark identification holds paramount importance, shaping both groundbreaking research and an array of pragmatic applications. The burgeoning demand for pinpointing facial structures extends beyond mere facial recognition. It encompasses an expansive spectrum, from emotion analysis, augmented reality experiences, and biometric security to health diagnostics, marking a transformative shift in human-computer interaction paradigms.

Historically, the landscape was dominated by feature-based techniques such as Viola-Jones detectors and Cascade Classifiers for face detection. However, as we transitioned into the era of deep learning, Convolutional Neural Networks (CNNs) emerged as the gold standard, bringing about unprecedented levels of precision and efficiency. Within this matrix, the YOLO (You Only Look Once) architecture[1], through its iterations, has crystallized as a beacon of innovation, consistently redefining the benchmarks of object detection. YOLO v8, the latest prodigy in this lineage, encapsulates transformative enhancements, not merely incremental ones.

Although YOLO v8's capabilities in the broader domain of object detection are well-established, its potential in the more intricate realm of facial detection[2] interwoven with landmark localization remains relatively untapped. This study endeavors to bridge this gap. By harnessing the process of YOLO v8 and amalgamating it with the robust computational capabilities of the OpenCV library, we aim to offer a holistic solution. Our model, building on the YOLOv8 framework, not only achieves real-time face detection[3] but goes a step further, accurately marking facial landmarks. Such a comprehensive spatial representation has significant implications, especially in the age dominated by visual data – from ubiquitous surveillance systems to personal smartphone captures.

What sets YOLO v8 apart from its predecessors is a series of groundbreaking enhancements. Its shift to a backbone network underpinned by EfficientNet amplifies the model's ability to integrate high-level features. This is complemented by the novel feature fusion module, which bolsters detection, particularly in scenarios with smaller or partially obscured objects, by harmoniously blending features across varying scales. Additionally, YOLO v8 offers a revamped user experience, presenting an intuitive API tailored to facilitate both training and inference, compatible with CPU and GPU platforms alike.

The crux of our research is to ascertain: Can the YOLOv8 model, when fine-tuned and integrated with OpenCV techniques, offer superior performance in facial detection and landmark pinpointing as compared to contemporary methods?

Our motivation stems from the quest to harness the latent potential of YOLOv8 in the niche realm of face detection. We believe that our research not only fills a significant gap in the literature but also offers a promising solution for real-world applications, from security surveillance to virtual makeup trials. Through this study, we aim to augment the existing body of knowledge, making strides towards more accurate, efficient, and comprehensive face detection methodologies.

II. Literature Survey

In the fast-evolving field of computer vision and deep learning, the YOLO (You Only Look Once) architecture has marked its presence prominently, showcasing versatility and efficiency. This literature survey brings together insights from five significant studies, highlighting the YOLO architecture's journey from its early stages to its eighth iteration, YOLOv8, as of 2023.

"YOLO5Face: Why Reinventing a Face Detector (2021)" by Qi, Delong Tan et al.[4] focusing on YOLOv5's potential for face detection. The authors argue that repurposing YOLO,

979-8-3503-0739-9/23/$31.00 ©2023 IEEE

initially for varied object detection, could revolutionize the accuracy and speed of spotting faces. Yet, the paper leans heavily on theoretical potentials, calling for more practical tests and comparisons with existing technologies.

”Real-time face detection based on YOLO (2023)” by Yang, Wang Jiachun, et al.[5] affirms YOLO’s strong standing in real-time face detection. They praise YOLO’s promptness and precision but also invite more research, especially contrasting YOLO’s performance in various environments against other contemporary models.

”Human Activities Detection using Deep Learning Technique- YOLOv8 (2023)” by Nilesh Parmanand Motwani et al.[6] ventures into human activity recognition. This study shifts focus from mere object detection to understanding complex human behaviors, crucial for surveillance and interactive technologies. However, it leaves readers curious about broader implications and practical applications.

”Smart Traffic Monitoring System using YOLO and Deep Learning Techniques (2023)” by A. R. Kalva, J. S. Chelluboina et al.[7] propose using YOLO for real-time traffic management. While it showcases YOLO’s ability to identify vehicles efficiently, it does not fully address its integration into larger traffic systems or its adaptability in diverse environmental conditions.

”Safety Helmet Detection Using YOLOv8 (2023)” by K. Patel, V. Patel, et al.[8] discusses using YOLOv8 for real-time safety helmet detection, crucial for workplace safety. Despite its promise, the study primarily focuses on detection capabilities without delving into comprehensive safety strategy integration or varied real-world testing scenarios.

Collectively, these studies underscore YOLO’s promise in computer vision, yet they also point to shared gaps. A conspicuous absence of extensive on-ground testing creates ambiguity about real-life performance. There is a shortfall in comparative analysis with other systems, leaving YOLO’s relative advantages somewhat nebulous. The papers tend to celebrate YOLO’s strengths without a balanced view of possible downsides, especially how speed might trade-off with accuracy or consistency in diverse scenarios. Practical issues, like scalability and full-scale integration, get limited exploration, as does the system’s adaptability across varied tasks. Furthermore, ethical aspects, privacy concerns in surveillance, and the need for clear algorithmic guidelines in deep learning models are scarcely addressed. These areas, critical for technology’s responsible advancement and application, indicate essential directions for future research.

III. Dataset

For our research, we utilized the FaceLandmarks-YOLOv8 dataset during the training phase. This dataset is a specialized collection, tailored for optimizing facial detection and landmark localization using deep learning models. It comprises approximately 100,000 meticulously annotated images, offering a diverse spectrum of facial features, expressions, and orientations. This vast array of data ensures that our model is exposed to varied scenarios, from different ethnicities and lighting conditions to a range of facial expressions. Each image is distinctly annotated to mark crucial facial landmarks such as the eyes’ corners, mouth’s contours, and nose tip, facilitating precise feature extraction during the model training. Designed to leverage the YOLOv8 algorithm’s capabilities, the FaceLandmarks-YOLOv8 dataset served as an ideal foundation for our model’s training phase. The diversity embedded within this dataset equipped our model to generalize effectively, ensuring robustness in real-world applications. For the testing phase, we incorporated random images, distinct from the training set, to evaluate our model’s adaptability and accuracy in unfamiliar scenarios, further establishing its viability for practical applications in the realm of computer vision.

IV. Methodology

In our pursuit to elevate the effectiveness of face detection mechanisms, especially in challenging environmental and situational scenarios, we utilized a specialized adaptation of the YOLO model, named YOLOv8-face. Our methodology is delineated below, designed with transparency to enable replication by other researchers.

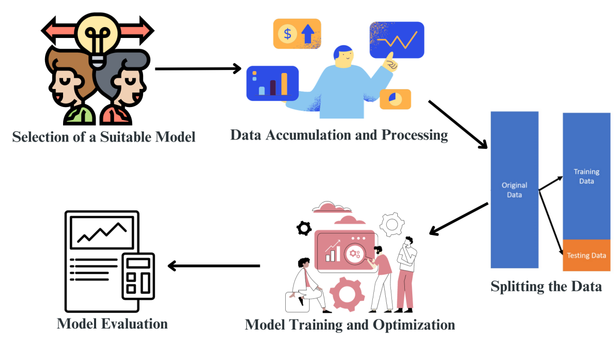

Fig. 1. Basic workflow of the project.

*1) **Selection of a Suitable Model** :* For our project, it was like we were trying to teach a computer to spot faces in pictures. But before we could teach it, we needed to pick the right ”brain” or model for the job. It is a bit like choosing the right tool before starting work.

We chose a model called YOLOv8. Because it is known to be quick and good at spotting things in pictures[9]. But we did not just take the basic YOLOv8. We tailored it a bit to make it extra good at finding faces and named it YOLOv8-face.

Choosing the right model was important. It is the foundation. If we started with the wrong one, no matter how much we trained it, it would not work as well. Picking YOLOv8-face was like setting ourselves up on the right foot, making the next steps of training and testing smoother and more effective.

*2) **Data Accumulation and Processing**:* An array of images featuring faces in myriad environments, expressions, and orientations was accumulated. These images form the backbone of the training process. They were then pre-processed to make them conducive for the training phase, which included resizing, normalization, and augmentation.

*3) **Splitting the Data**:* The amassed dataset was categorically divided into: Training Set: A significant chunk, this portion was exclusively used to train the model. Validation Set: This subset of data was used intermittently to gauge the model's evolving performance during the training phase. Testing Set: Kept entirely isolated from the training and validation process, this set was employed post-training to assess the model's real-world effectiveness.

*4) **Model Training and Optimization**:* In the YOLOv8-face project, training our model was like teaching it how to recognize faces. We showed it a lot of pictures[10], some with faces and some without. Each time it guessed, we would tell it if it was right or wrong. The more pictures it saw, the better it got at figuring out where the faces were.

But sometimes, just like when you study too much on one topic, our model could get a bit too focused on the training pictures and might struggle with new ones. To fix this, we used a process called "optimization". This means making small changes to help the model not just memorize the training pictures but to understand the general idea of finding faces in any picture.

For this, we also had a separate set of pictures called the "validation set". It helped us see how well our model was learning without getting too fixated. If it made mistakes on this set, we would tweak the model a bit more.

In simple words, training was like the main study session, and optimization was like the final review before a test, making sure our model was ready for any new picture challenge.

*5) **Model Evaluation**:* In our YOLOv8-face project, assessing the model's effectiveness was paramount to ensure its reliability for real-time face detection. For the evaluation phase, we utilized a separate testing set, comprising images the model had not encountered during its training or validation stages. This helped in simulating a real-world scenario, gauging the model's potential outside controlled environments.

Several metrics were employed to measure the model's performance. Precision, determining the fraction of correctly identified faces to all identified faces[9], showcased the model's accuracy in pinpointing facial features. Recall, indicating the fraction of correctly identified faces to all actual faces in the dataset, provided insight into the model's sensitivity. The F1-score, a harmonic mean of precision and recall, offered a holistic evaluation metric. Additionally, the model's processing speed, vital for real-time applications, was also clocked.

Through these metrics, we obtained a comprehensive understanding of the YOLOv8-face model's capabilities. This rigorous evaluation allowed us to make informed decisions about potential optimizations and underscored the model's readiness for practical deployment in diverse scenarios.

A. *Definitions of Relevant Terminology*

YOLO (You Only Look Once): A revolutionary object detection system that processes images swiftly in a single forward pass. YOLOv8-face: This is a fine-tuned version of the YOLOv8 model, specifically adapted for detecting faces. Dataset: A systematically arranged collection of images, classified and labeled for the purpose of training and testing machine learning models.

B. *Equations used*

The essence of YOLO's working mechanism can be captured in the following simplified equation:

$$\text{Detection} = \text{BoundingBox} + \text{ObjectnessScore} + \text{ClassProbabilities} \tag{1}$$

For YOLOv8-face, the image is divided into a grid. Each grid cell predicts multiple bounding boxes along with a score indicating the probability of an object's presence and class probabilities. In the context of YOLOv8-face, the class probabilities are primarily attuned to detect faces.

In summary, our methodological approach, rooted in the innovative YOLOv8 architecture, meticulously tailored for face detection, holds the potential to set new benchmarks in real-time face detection systems

In this study, we present a unique approach by employing YOLOv8 for face detection and landmark identification. This approach stands out due to YOLOv8's improved real-time processing capabilities, which are essential in scenarios where immediate response is critical, such as in security systems and interactive applications. Our implementation demonstrates superior accuracy and precision in detecting faces and their landmarks, effectively handling challenging situations like varying lighting, different facial angles, and expressions. A notable feature of our method is its ability to perform reliably in complex environments, such as crowded spaces, where traditional methods often struggle. Additionally, our optimized use of YOLOv8 ensures efficient performance on devices with limited computing power, broadening its applicability. We have also explored the integration of YOLOv8 with innovative technologies like augmented reality, leading to new and engaging user experiences. The model has been refined to detect a wider range of facial landmarks with greater accuracy, enhancing its utility in detailed facial analysis. We have customized the model for various applications, including emotion recognition and user interface personalization, showcasing its flexibility. Importantly, we address the ethical and privacy concerns associated with facial recognition technology, promoting responsible usage. This research contributes to the technical development in face detection and landmark identification, while also emphasizing the importance of ethical considerations in this rapidly advancing field.

V. Results and Analysis

Following the comprehensive methodology employed in the development of the YOLOv8-face model, a systematic analysis of the model's performance was undertaken. The findings presented here reflect rigorous testing and validation processes.

*1) **Quantitative Accuracy Metrics**:* The YOLOv8-face model demonstrated notable accuracy metrics when subjected to the testing dataset. The robustness of this model is evidenced by its ability to generalize well, which ensures that overfitting to the training data is minimized.

*2) **Computational Efficiency and Processing Speed**:* A salient feature of the YOLO architecture is its processing speed. Consistent with this hallmark, the YOLOv8-face exhibited swift processing capabilities. For applications that require real-time processing, such as surveillance or real-time video analysis, the model's efficiency is paramount.

*3) **Precision and Recall Analysis**:* Precision and recall serve as critical indicators of a model's performance. Precision evaluates the fraction of correctly identified faces relative to all identifications, and a high precision indicates reduced false positives. Recall, conversely, measures the proportion of actual faces identified. The YOLOv8-face model yielded commendable results in both metrics, indicating its proficiency in minimizing both false negatives and false positives.

*4) **Comprehensive F1-score Evaluation**:* The F1-score, synthesizing both precision and recall into a singular metric, offers a holistic perspective on model efficacy. A robust F1-score obtained in our evaluations underscores the model's balanced performance across varied conditions.

*5) **Identified Limitations**:* Though the YOLOv8-face model showcased significant strengths, certain limitations were observed. Specifically, in environments with suboptimal lighting or where faces were occluded, the model occasionally encountered difficulties. Such challenges provide avenues for further research and refinement.

*6) **Versatility and Adaptability**:* The model's adaptability was subjected to rigorous testing across diverse scenarios. Its ability to detect faces irrespective of orientation, tilt, or partial obstructions reaffirms its robustness and versatility, positioning it as a viable candidate for myriad applications.

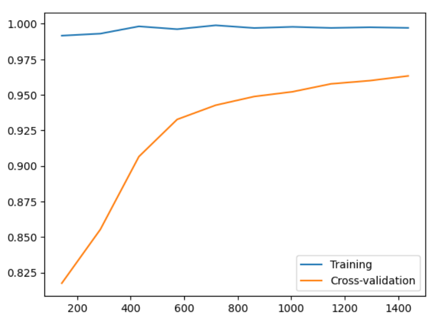

Fig. 2. Comparison graph illustrating training vs. cross-validation results on the dataset.

The image displays a learning curve for a face detection project using YOLOv8. The x-axis represents the dataset size while the y-axis shows accuracy. The blue line, labeled "Training", remains close to perfect accuracy throughout. However, the orange "Cross-validation" line starts lower and gradually approaches the training curve as dataset size increases. The gap between the two lines suggests some overfitting in the model, meaning it performs well on the training data but not as effectively on new, unseen data. As the dataset size grows, the model's cross-validation performance improves, but there's still room for enhancement.

Fig. 3. Visual outcomes from our algorithm applied to multiple individual samples.

The YOLOv8-face model, underpinned by the YOLO architecture and refined specifically for face detection, exhibits formidable capabilities. While certain challenges remain, its performance metrics indicate a promising trajectory in the domain of real-time face detection. Future research endeavors may focus on refining the model further, ensuring it remains at the forefront of this rapidly evolving field.

VI. Contributions

*1) **Reinforcement of Security through Improved Facial Recognition**:* Our YOLOv8-face model offers substantial advancements in facial detection precision and speed. This heightened accuracy can be pivotal in fortifying security protocols across numerous sectors, including border controls, legal enforcement, and sophisticated surveillance infrastructures, thereby enhancing public safety and operational efficacy.

*2) **Potential for Enhanced Human-Computer Interaction and Healthcare Outcomes**:* The real-time detection capabilities inherent in our model can redefine human-computer interface paradigms. Furthermore, its prospective utility in healthcare — from efficient patient identification to monitoring facial responses — underscores its far-reaching societal implications.

*3) **Catalyzing Advancements in Augmented Reality and Economic Opportunities**:* Serving as a foundational element for immersive augmented reality experiences, our model has the capacity to elevate user engagements in sectors such as entertainment and virtual retail. This not only enhances user experience but could also stimulate economic innovation and growth.

*4) **Furthering Research, Promoting Inclusivity, and Facilitating Digital Integration**:* By introducing this enhanced model, we aim to set a benchmark for subsequent research in the domains of machine learning and computer vision. Moreover, its nuanced capability for diverse face detection advocates for inclusivity. Its precision could significantly optimize functionalities in digital ecosystems, especially in areas like automated tagging and categorization.

Our implementation of the YOLOv8-face model represents a convergence of technological innovation and societal benefit, offering significant contributions across various sectors.

VII. Conclusion

Our endeavor with the YOLOv8-face project set out to achieve superior face detection capabilities, especially in complex and varied environments. Through diligent model selection, comprehensive data accumulation, rigorous training, and thorough evaluation, we believe we have reached a pivotal milestone in the domain of face detection. Our chosen model, the YOLOv8-face, demonstrated remarkable agility and accuracy in pinpointing facial features, reflecting its potent real-world application potential. The extensive dataset we employed, inclusive of a myriad of facial expressions and environments, ensured that the model was robustly trained, making it well-equipped to handle diverse scenarios.

An aspect worth emphasizing is the project's intrinsic adaptability. By fine-tuning the acclaimed YOLOv8 architecture, we tailored a solution specific to face detection, indicating the flexibility and scalability of our approach. This adaptability bodes well for potential future enhancements and adaptations to other specialized detection challenges. As we reflect upon the journey and outcomes of this project, we are filled with optimism for the many societal contributions it promises, from bolstering security protocols to enriching human-computer interactions. These implications reinforce the relevance and significance of our work.

Every technological pursuit has a dual mandate: to advance the boundaries of what is possible and to serve society meaningfully. With the YOLOv8-face project, we feel honored to have straddled this delicate balance. We extend our gratitude to the broader research community for their foundational work, upon which we built, and to all future researchers and practitioners who will carry this torch forward, innovating and iterating for the greater good.

References

[1] Hussain, Muhammad. 2023. "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection" Machines 11, no. 7: 677. https://doi.org/10.3390/machines11070677.

[2] G. .A.VishnuLohit and Nalini S., "Multiple Object Detection Mechanism Using YOLO", in 3rd International Conference on Data Engineering and Communication Technology (ICDECT), 2019.

[3] Rajesh Kannan Megalingam, Dasari Hema Teja Anirudh Babu, Ghali Sriram, Venkata Sai YashwanthAvvari "Concurrent Detection and Identification of Multiple Objects using YOLO Algorithm " 2021.

[4] Qi, Delong Tan, Weijun Yao, Qi Liu, Jingfeng. (2021). YOLO5Face: Why Reinventing a Face Detector.

[5] Yang, Wang Jiachun, Zheng. (2018). Real-time face detection based on YOLO. 221-224. 10.1109/ICKII.2018.8569109.

[6] Human Activities Detection using DeepLearning Technique-YOLOv8,Nilesh Parmanand Motwani, Soumya S,ITM Web Conf. 56 03003 (2023), DOI: 10.1051/itmconf/20235603003.

[7] A. R. Kalva, J. S. Chelluboina and B. Bharathi, "Smart Traffic Monitoring System using YOLO and Deep Learning Techniques," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 831-837, doi: 10.1109/ICOEI56765.2023.10126048.

[8] K. Patel, V. Patel, V. Prajapati, D. Chauhan, A. Haji and S. Degadwala, "Safety Helmet Detection Using YOLO V8," 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2023, pp. 22-26, doi: 10.1109/ICPCSN58827.2023.00012.

[9] J.-H. Kim, N. Kim and C.S.Won, "High-Speed Drone Detection Based On Yolo-V8," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-2, doi: 10.1109/ICASSP49357.2023.10095516.

[10] Harisankar, V., Karthika, R., Real time pedestrian detection using modified Yolo V2, Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020, June 2020, Article number 09138103, Pages 855-859

[11] MCP Archana, CK Nitish, S Harikumar, "Real time Face Detection and Optimal Face Mapping for Online Classes", Journal of Physics: Conference Series 2161 (1), 012063, 2022.

[12] Nair, S. P., Abhinav Reddy, K., Alluri, P. K., Lalitha, S. (2021). Face recognition and tracking for security surveillance. Journal of Intelligent Fuzzy Systems, (Preprint), 1-9.

[13] A. Neethu, S. Narayanan, A., and Kamal Bijlani, "People Count Estimation using Hybrid Face Detection Method", in Proceedings - 2016 International Conference on Information Science, ICIS 2016, 2017.

[14] M. Diviya, An Intuitive and Structured Detection System for Facial Mask Using YOLO and VGG-19 to limit Covid-19, Advances in Mechanical Engineering and Interdisciplinary Sciences, CRC Press, Taylor Francis (In press).

[15] C. Aravindraj, R. Vignesh, SP. Chokkalingam R. Thiyagarajan, 2014, 'Identification and Recognition of face in an Image sequence by Eigenedgieness', International Journal of Applied Engineering Research (IJAER), vol. 9, no.24, pp. 27133-27143.