COMPARATIVE EVALUATION OF CLASSICAL AND DEEP LEARNING FACE DETECTORS ON FDDB, WIDER FACE, AND CELEBA DATASETS

Bengisu Adsız
Department of Computer Engineering
Hacettepe University
bengisuadsiz@hacettepe.edu.tr

ABSTRACT

This report presents a comparative study of classical and modern deep learning-based face detection models. This project evaluates the performance of the traditional Viola-Jones Haar Cascade classifier and two state-of-the-art deep learning models, YOLOv8n-Face and SCRFD, on benchmark datasets: FDDB, WIDER FACE, and CELEBA. The models are compared using standard detection metrics under different configurations. This project provides detailed analysis, implementation challenges, and discuss improvement strategies. This study aims to understand how well legacy methods hold up against current state-of-the-art approaches in varied face detection conditions.

1 INTRODUCTION

Face detection remains a core component of numerous real-world applications, including surveillance systems, user authentication, photo tagging, and human-computer interaction. The primary motivation for this project comes from a curiosity to understand the evolution of face detection models—starting from classical rule-based techniques (Viola-Jones Haar Cascades) to modern deep learning approaches (YOLOv8n-Face, SCRFD).

This report aims to perform a structured comparison of three different paradigms of face detection:

- The **Viola-Jones** framework with Haar Cascades: a classical yet historically significant and computationally efficient method.
- **YOLOv8n-Face**: a modern, anchor-free real-time face detector based on a lightweight YOLOv8 architecture.
- **SCRFD**: abbreviation of ”Sample and Computation Redistribution for Face Detection” a fast and compact deep learning model optimized for mobile deployment and small face detection.

Through extensive experimentation, this study investigates how traditional approaches perform in contrast to deep learning-based models under real-world variability, such as scale, occlusion, and pose. The experiments provided insights into the strengths and weaknesses of each approach and demonstrated the dramatic improvements in face detection achievable through deep learning.

2 MODEL DESCRIPTION

2.1 VIOLA-JONES HAAR CASCADES

The Viola-Jones face detector was the first robust real-time face detection algorithm and remains foundational in computer vision literature. It operates through the following components:

- **Haar-like Features**: Simple rectangular patterns that can capture edges, lines, and intensity differences. Each feature computes the difference in pixel sums between adjacent rectangular regions.

- **Integral Image Representation**: A data structure that enables rapid computation of these features in constant time regardless of window size.
- **Adaboost-based Feature Selection**: From an over-complete feature pool, Adaboost is used to select the most informative ones to train weak classifiers.
- **Cascade Architecture**: Classifiers are arranged in a cascade so that non-face windows are quickly rejected at early stages, allowing only promising regions to proceed further in the cascade.

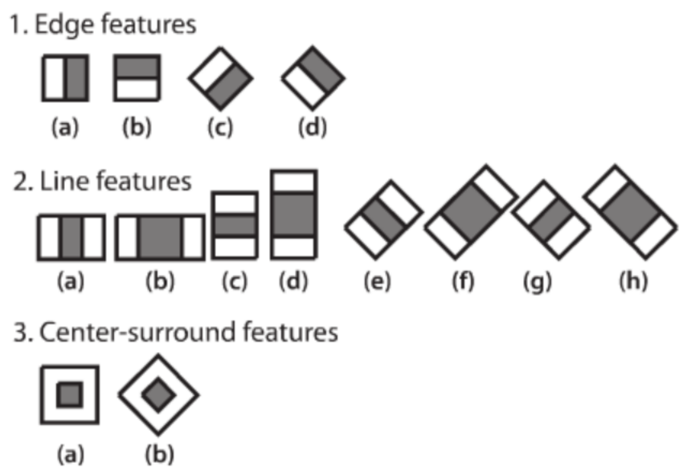

Figure 1: Viola-Jones face detection pipeline overview

Viola-Jones is known for its simplicity and low computational cost, making it suitable for embedded systems or constrained environments. However, its reliance on fixed geometric templates and handcrafted features limits its robustness against challenging face variations in pose, illumination, and size.

2.2 YOLOv8n-Face

YOLOv8n-Face builds upon the YOLOv8 architecture, which uses an anchor-free detection mechanism for efficiency and flexibility. The key structural components are:

- **Backbone (C2f blocks)**: A lightweight CSP-based network extracts hierarchical feature maps at multiple scales. C2f blocks improve information flow while keeping the model size compact.
- **Neck (Feature Aggregation)**: Typically includes a PAN (Path Aggregation Network) or BiFPN-like module that fuses features across different scales to improve detection of objects with varied sizes.
- **Detection Head**: Outputs bounding box coordinates, confidence scores, and optional class probabilities using a decoupled head architecture. The anchor-free format simplifies training and reduces tuning effort.
- **Post-processing**: Non-Maximum Suppression (NMS) is applied to remove overlapping predictions.

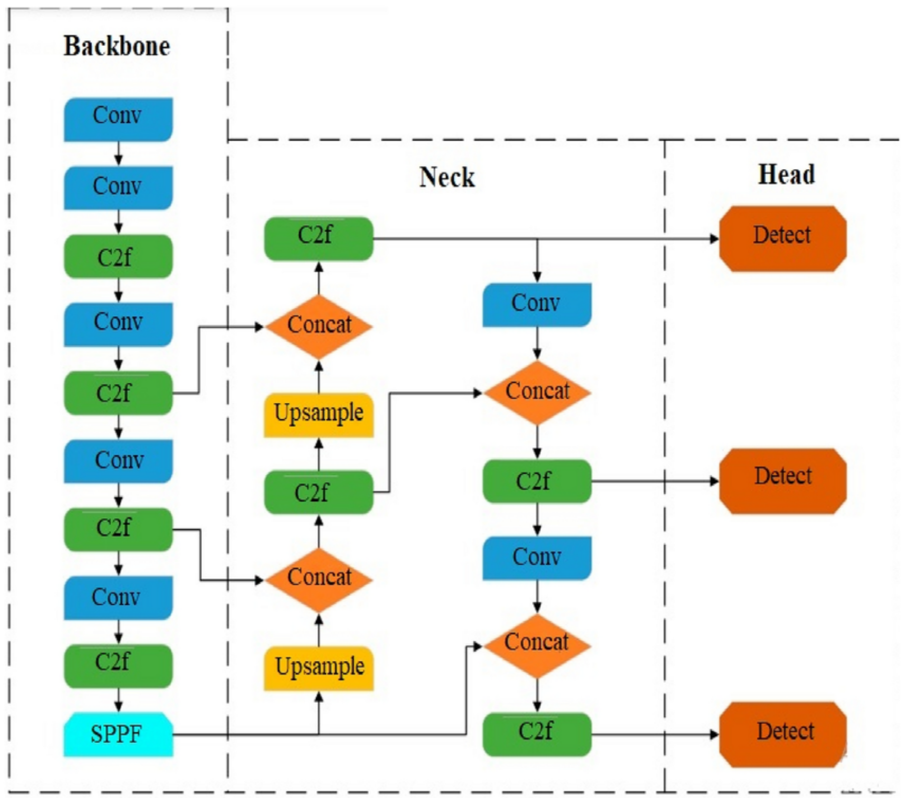

Figure 2: YOLOv8n-Face architecture with backbone, neck, and detection head

YOLOv8n-Face is especially powerful due to its balance between speed and accuracy, achieving real-time inference speeds while still maintaining robustness across detection scenarios.

2.3 SCRFD (INSIGHTFACE)

SCRFD is a high-efficiency face detector proposed by InsightFace. It is specifically designed to detect small faces under real-world constraints. The architecture includes:

- **Re-parameterized Backbone**: Uses a MobileNet-like lightweight structure with re-parameterization during inference to merge BatchNorm and Conv layers for latency reduction.
- **Detection Head**: Employs a dense anchor-based strategy. It outputs classification scores, bounding box regressions, and IoU estimations, allowing better filtering during post-processing.
- **Training Techniques**: The model is trained with OHEM (Online Hard Example Mining), GIoU loss, and multi-scale face augmentations using the WIDER FACE dataset.

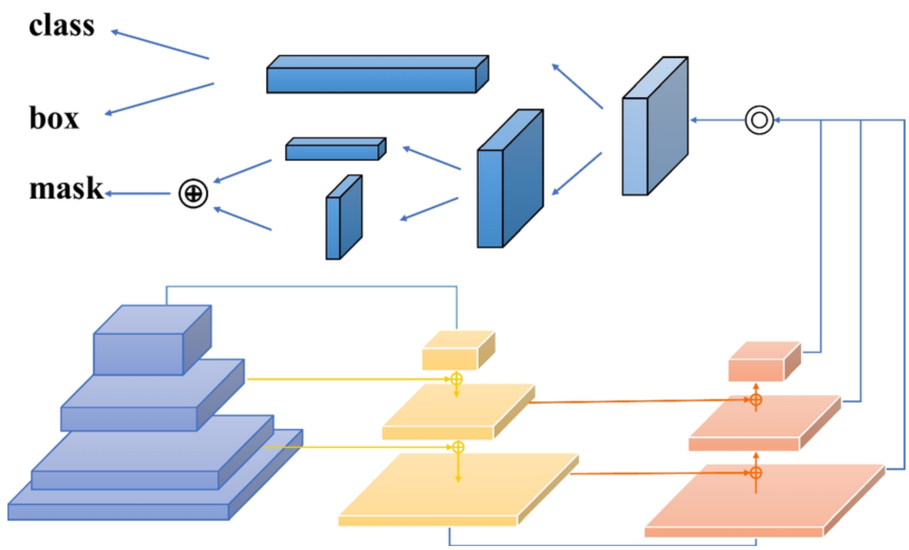

Figure 3: SCRFD model architecture showing backbone and detection head

SCRFD is optimized for edge devices and delivers strong performance in detecting tiny, occluded, or rotated faces.

3 Experimental Setup

All experiments were conducted using the Google Colab environment, which provided a flexible and GPU-accelerated workspace ideal for running deep learning workloads. The runtime environment was configured with an NVIDIA A100 GPU, offering high-speed computation that significantly reduced model inference and evaluation times.

The implementation utilized Python 3 and libraries such as PyTorch for neural network operations, OpenCV for image handling, and ONNXRuntime for SCRFD model inference. Datasets (FDDB, WIDER FACE, and CELEBA) were loaded either from Kaggle datasets or uploaded directly into Colab's working directory.

4 Experimental Results and Comparisons

4.1 FDDB Dataset

Table 1: FDDB Dataset - Face Detection Performance

Method	**TP**	**FP**	**FN**	**Precision**	**Recall**	**F1 Score**
Haar (No NMS)	3820	63912	1351	0.056	0.739	0.105
Haar (With NMS)	3417	32461	1754	0.095	0.661	0.166
YOLOv8 (Conf=0.3)	2498	44863	2673	0.053	0.483	0.095
YOLOv8 (Conf=0.4)	2483	43735	2688	0.054	0.480	0.097
YOLOv8 (Conf=0.2, NMS)	2576	49239	2595	0.050	0.498	0.090
SCRFD (IOU=0.3)	7437	42568	33165	0.149	0.183	0.164

4.2 WIDER FACE Dataset

Table 2: WIDER FACE Dataset - Face Detection Performance

Method	TP	FP	FN	Precision	Recall	F1 Score
Haar (No NMS)	5386	3586	23203	0.600	0.188	0.287
Haar (With NMS + Side)	2819	1888	23068	0.599	0.109	0.184
YOLOv8 (Conf=0.1)	20059	21501	8530	0.483	0.702	0.572
YOLOv8 (Conf=0.15)	19064	12479	9525	0.604	0.667	0.634
YOLOv8 (Conf=0.2)	18016	7629	10573	0.703	0.630	0.664
YOLOv8 (Conf=0.2, NMS)	17853	5716	10736	0.757	0.624	0.685
SCRFD (IOU=0.3)	2046	18819	17715	0.098	0.104	0.101

4.3 CELEBA Dataset

Table 3: CELEBA Dataset - Face Detection Performance

Method	TP	FP	FN	Precision	Recall	F1 Score
Haar (No NMS)	99	2865	2901	0.033	0.033	0.033
Haar (With NMS + Side)	97	2870	2903	0.033	0.032	0.033
YOLOv8 (Conf=0.1)	305	2832	2695	0.097	0.102	0.099
YOLOv8 (Conf=0.15)	305	2795	2695	0.098	0.102	0.100
YOLOv8 (Conf=0.2)	305	2770	2695	0.099	0.102	0.100
YOLOv8 (Conf=0.2, NMS)	305	2748	2695	0.100	0.102	0.101
SCRFD (IOU=0.3)	–	–	–	–	–	–

5 Observations and Achievements

To provide a clear comparison of model performances, I generated a visual summary of F1 Scores across the three benchmark datasets.

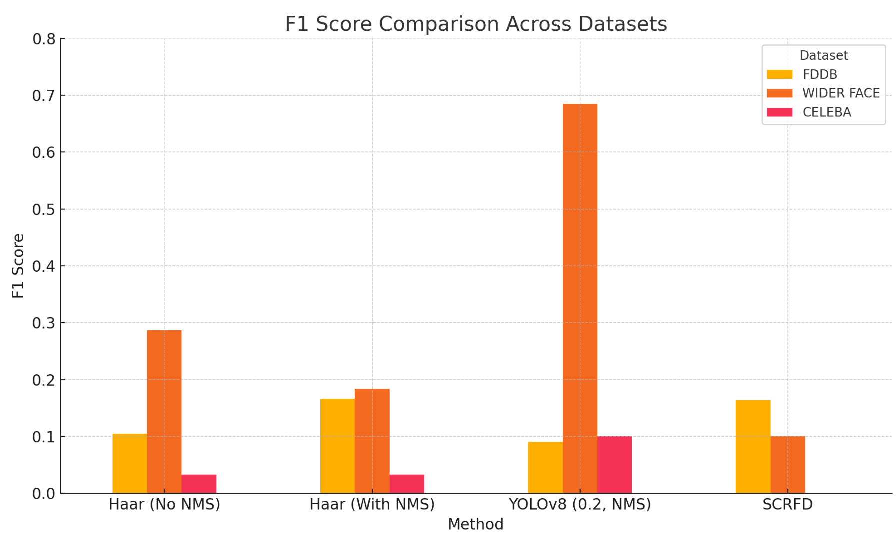

Figure 4: F1 Score comparison of face detection methods across FDDB, WIDER FACE, and CELEBA datasets.

From the results, we can make the following key observations:

- **Haar Cascades**, while historically significant, showed limited applicability to modern, unconstrained datasets. On FDDB, it maintained relatively high recall but suffered from extremely high false positives. On WIDER FACE and CELEBA, it showed diminished performance due to the complexity and variability of faces.
- **YOLOv8n-Face** consistently outperformed the classical method in terms of F1 Score across datasets. Its real-time capability and anchor-free detection mechanism make it highly suitable for practical applications. Fine-tuning confidence thresholds and applying NMS helped to further improve performance.
- **SCRFD**, although designed for small and occluded faces, yielded mixed results. It performed well on FDDB, indicating robustness on frontal faces. However, it struggled with WIDER FACE due to its higher sensitivity to background clutter and lack of additional fine-tuning.

Through these experiments, I was able to:

- Quantitatively assess the limitations of classical methods on modern, challenging datasets.
- Verify that while Haar Cascades can still achieve high recall on FDDB, their high false positive rate makes them impractical without careful post-processing.
- Show how YOLOv8n-Face can outperform older methods significantly in both precision and F1 score, especially when hyperparameters like confidence threshold and NMS are properly tuned.

These experiments reinforced my understanding of face detection techniques and gave me hands-on experience in benchmarking, tuning, and comparing models under fair conditions.

6 Challenges and Discussions

Throughout this study, multiple implementation and evaluation challenges were encountered:

- **Haar Cascade Models**: Implementing and evaluating the Haar Cascade with profile detection required additional logic to combine detections and apply NMS. Haar detectors produced excessive false positives, especially on the WIDER FACE dataset.
- **YOLOv8-Face Setup**: Setting up YOLOv8-Face in Colab required precise control over confidence, IoU, and minimum box size filters to stabilize performance.Nevertheless, tuning the model provided valuable insights into object detection hyperparameters.
- **SCRFD ONNX Evaluation**: The SCRFD model required ONNX runtime support and used custom preprocessing and output decoding. Evaluation on datasets such as FDDB and WIDER FACE demanded extra steps to ensure bounding box compatibility.
- **CELEBA Format Issue**: None of the models could achieve reliable performance on CELEBA due to a lack of official bounding box annotations. Manual approximations failed to reflect true ground truth, leading to unfairly low precision and recall values. This limitation particularly impacted SCRFD, where no valid detections were generated due to mismatched ground truth format.

Future directions may involve:

- Training models on a subset of CELEBA with custom bounding box annotations.
- Applying data augmentation for better generalization.
- Exploring hybrid models that combine the speed of YOLO with the structural insights of anchor-based methods like SCRFD.

Acknowledgments

I would like to thank the instructor of CMP719A for guidance and providing resources to explore modern face detection techniques.

REFERENCES

[1] Viola, P. & Jones, M. (2001). *Rapid Object Detection*. CVPR.

[2] Yisihak, H.M. and et al. (2024). *Advanced Face Detection with YOLOv8: Implementation and Integration into AI Modules*. Open Access Library Journal,.

[3] Jianzhu Guo et al. "Sample and Computation Redistribution for Efficient Face Detection.", (2021).

[4] *Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, (2016)*

[5] Wang, G. et al. (2023). *EfficientFace: An Efficient Deep Network with Feature Enhancement for Accurate Face Detection*. CVPR.

[6] YOLOv8 Documentation. Ultralytics.

CODE REPOSITORY

The complete and well-documented codebase for this project is available at:
`https://github.com/tatlicadikiki/CMP719A-FACE-DETECTION-MODEL-COMPARISON`