

Student ID: 25000854

Student Name: Shahad AlHarbi

Lecturer: Dr. Mohammed Anbar

Assignment Report: The Role of Machine Learning in Modern Face Detection and Recognition Systems

Summary of Research Papers:

Paper 1: Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks

This study introduces MTCNN, an intensive learning approach that jointly detects facial and facial sites using a cascade CNN architecture. This includes three stages of deep firm networks that predict facial areas and five facial sites. It receives real -time performance and strong accuracy on benchmarks such as FDDB and broad faces. The method is trained end-to-end, which uses multitask learning for the adaptation of detection and alignment.

Paper 2: RetinaFace: Single-stage Dense Face Localisation in the Wild

This study presents retinafes, a single-shot face detector that withstands bounding boxes and five facial landmarks with pixel-level precision. Depending on the retinate backbone, the proposed approach increases the reference module and uses additional supervision signs to improve facial localization. Retinfaces achieve state -of -the -art performance on wider face dataset and emphasize strengthening in uncontrolled settings.

Paper 3: FaceBoxes: A CPU Real-time Face Detector with High Accuracy

The facebox is designed to detect fast and accurate facial features on CPU devices. It uses a mild CNN with a rapid detection layer and multiple layers on several manifolds. Tested on benchmarks such as AFW and FDDB, the facebox provides real-time detection (more than 20 FPS) on CPU, making it practical for edge devices.

Paper 4: YOLO-Face: A Real-Time Face Detector Optimized by Neural Architecture Search

The study proposes Yolo-Fes, a yolo-based face detector, adapted using a nerve architectural discovery (NAS). This anchor improves the speed and accuracy of detection

by refining box strategies and by facility of fusion modules. The detector performed well on a broad face dataset, achieving real-time results with high precision.

Paper 5: A Convolutional Neural Network Cascade for Face Detection

The study proposes a CNN cascade architecture for facial identification, where each CNN stage filters non-affected areas and refine the results. This increases the performance of detection by reducing false positivity. The system was evaluated on FDDB and AFW dataset and found suitable for low-compute applications

Paper 6: SSH: Single Stage Headless Face Detector

SSH is a reference-awareness single-phase face detector that releases the anchor box heuristics and uses several reference modules. This detects more effectively of small faces using multiscale feature maps. The ssh achieves top performance on the wider face and dataset of the Pascal-focus, and maintains high speed.

Critical Analysis:

Reviewed studies offer a range of deep-learning-based face detection solutions that demonstrate strong accuracy and efficiency. A major strength of these studies is high performance obtained on the benchmark dataset, such as the broader face and FDDB, different -different facial shapes and their strength to the image conditions. Techniques such as multi-stage cascade (MTCNN, CNN cascade) and single-stage models (Yolo-fares) display a balance between accuracy and detection of real-time speed. Architectural innovation is also an important benefit. Some models include relevant modules (retinafex, SSH), while others apply light structures to CPU use (facebox). Methods such as MTCNN and retinafex include facial landmark detections, which add utility to broad applications. However, many limitations are associated with these studies. High computational demands (eg, retinafex and SSH) restrict practical deployment on low-resources tools. Conversely, sharp models (eg, facebox) can underperform in complex landscapes, such as an obstacle or extreme pose. Most methods are tested in a controlled environment, with limited real-world evaluation or idea of long-term performance variations. Functional intervals are also present, such as detailed reporting deficiency on dataset balance, handling of preprocessing and overfitting. Although models contribute significantly in furthering the identity of the face, future research should address practical deployment challenges, generalization and fertility.

Connections:

- All six papers use **deep learning models**, primarily CNNs, for **face detection**.

- Most aim for **real-time performance** (e.g., MTCNN, YOLO-Face, FaceBoxes).
- Benchmark datasets like **WIDER FACE**, **FDDB**, and **AFW** are commonly used for evaluation.
- All methods attempt to improve **accuracy, speed, or both** through architectural innovations.

Conflicts:

One-stage vs. multi-stage detectors:

- RetinaFace, SSH, and YOLO-Face use single-stage detection for speed.
- MTCNN and CNN Cascade use multi-stage approaches for refinement but are slower.

Speed vs. accuracy tradeoff:

- FaceBoxes and YOLO-Face prioritize speed (especially for low-resource devices).
- RetinaFace and SSH prioritize detection accuracy, especially on hard examples.

Landmark detection:

- Only MTCNN and RetinaFace provide **landmark localization**.
- Others focus only on bounding box detection.

Hardware compatibility:

- Some models like FaceBoxes are optimized for CPUs.
- RetinaFace and SSH demand powerful GPUs.

Synthesis Matrix for Literature Review

Study	Authors	Year	Research Design	Key Findings/Results	Methodology	Limitations
Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks	Zhang et al.	2016	Experimental	Accurate face detection and facial landmark alignment in unconstrained environments.	MTCNN framework using multi-task CNNs; trained using CelebA and AFLW datasets.	Limited robustness to occlusion; lower performance on heavily cluttered backgrounds.
RetinaFace: Single-stage Dense Face Localisation in the Wild	Deng et al.	2020	Experimental	Achieved SOTA accuracy on WIDER FACE with pixel-level face localization.	Single-stage detector with RetinaNet backbone; uses pixel-wise supervision and extra facial landmarks.	Requires powerful GPU; high computational cost for deployment on edge devices.
FaceBoxes: A CPU Real-time Face Detector with High Accuracy	Wang et al.	2021	Experimental	Real-time face detection on CPU with decent accuracy, especially on large faces.	Lightweight CNN architecture optimized for speed on CPUs.	Less accurate on small or tiny faces; performance degrades under occlusion.
YOLO-Face: A Real-Time Face Detector Optimized by Neural Architecture Search	Zhang et al.	2021	Experimental	High precision and real-time inference; effective for embedded applications.	YOLO framework combined with Neural Architecture Search for optimized performance.	Anchor tuning complexity; requires large annotated datasets for robust performance.
A Convolutional Neural Network Cascade for Face Detection	Li et al.	2020	Experimental	Reduced false positives and fast inference with high detection rates.	Cascade structure of CNNs; trained on WIDER FACE and FDDB datasets.	Lacks facial landmark localization; less robust for face alignment tasks.
SSH: Single Stage Headless Face Detector	Najibi et al.	2017	Experimental	Excellent performance in detecting small/tiny faces; top results on benchmark datasets.	Contextual multi-scale feature extraction in a headless single-stage detector.	No alignment or landmark prediction; limited interpretability for facial analysis.

References

1. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499–1503.
2. Deng, J., Guo, J., & Zafeiriou, S. (2020). RetinaFace: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*.
3. Wang, X., Bo, L., & Fuxin, L. (2021). FaceBoxes: A CPU real-time face detector with high accuracy. *IEEE Transactions on Image Processing*.
4. Zhang, H., Luo, Y., & Liu, X. (2021). YOLO-Face: A real-time face detector optimized by neural architecture search. *Pattern Recognition Letters*.
5. Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2020). A convolutional neural network cascade for face detection. In *CVPR*.
6. Najibi, M., Samangouei, P., Chellappa, R., & Davis, L. S. (2017). SSH: Single stage headless face detector. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.