Report for Computer Vision Class Project: Cascading Facial Verification Models

Mehul, Gautam, Aditya Bagri, Aryan Jain

Placeholder Affiliation/Email

April 24, 2025

1 Problem Statement

Facial verification involves determining whether two face images belong to the same person, a task critical in applications like security systems, access control, and social media identity verification. In many real-world scenarios, the base rate—the probability that two faces match—is very low, often as low as 1%. In such cases, traditional accuracy metrics are inadequate because a model could achieve high accuracy by simply predicting all pairs as non-matches, failing to identify true matches effectively. To address this, we focus on the F1 score, which balances precision and recall, making it suitable for imbalanced datasets where matches are rare [1].

Our project introduces a novel Cascading Adaptive Face Verification (CAFV) framework, designed to be practical and efficient in low base rate scenarios. CAFV employs a sequence of models with increasing computational complexity: a lightweight model quickly filters out obvious non-matches, and only pairs with a high likelihood of matching proceed to a heavier, more accurate model. This approach aims to optimize both computational efficiency (measured in Mflops) and verification performance, particularly when the probability of a match is low.

We use the Labeled Faces in the Wild (LFW) dataset [2], a standard benchmark for face verification, which provides pairs of face images labeled as same or different. By subsampling LFW, we simulate base rates of 1%, 5%, 10%, and 50% to evaluate CAFV's performance across various scenarios.

Motivation

Low base rate scenarios are prevalent in practical applications. For example, in an access control system, most verification attempts come from unauthorized individuals, resulting in a low base rate of matches. Similarly, in social media, verifying identities against a small set of known users involves many non-matching pairs. Traditional face verification methods often assume a balanced distribution (e.g., 50% base rate), which is unrealistic in these contexts. CAFV addresses this by adapting its computational strategy to the base rate, ensuring high performance with minimal resource use.

Scope

The project focuses on facial verification using LFW [2], emphasizing low base rate scenarios. We will simulate CAFV's performance, compare it with baseline models, and analyze its efficiency in terms of F1 score and Mflops. The approach is designed for edge devices, where computational resources are limited [3], making Mflops a critical metric.

2 Related Work and Baselines

2.1 Broad Categories of Works

Face recognition research has advanced significantly, particularly in developing lightweight models for edge devices. Key categories relevant to our project include:

- **Lightweight Face Recognition Models:** These models prioritize computational efficiency while maintaining competitive accuracy. Techniques like depthwise separable convolutions, channel shuffling, and mixed precision reduce Mflops, making models suitable for resource-constrained environments. Examples include MobileFaceNet [4], ShuffleFaceNet [5], and PocketNet [6]. Other techniques include model compression [7] and efficient convolution operations [8].

- **Cascading Classifiers:** Inspired by methods like Viola-Jones for face detection [9], cascading in face verification uses a sequence of models with increasing complexity. Early stages reject non-matches quickly, while later stages provide accurate verification, optimizing computational efficiency.

- **Handling Imbalanced Datasets:** In low base rate scenarios, metrics like F1 score, precision, and recall are more informative than accuracy. Research in this area explores strategies to improve performance when positive classes (matches) are rare [1]. Adaptive methods [10] also consider variations in facial appearance.

2.2 Baselines

We select three baseline models with approximately 500 Mflops, representing lightweight yet effective face recognition architectures:

- **MobileFaceNets** (439.8 Mflops) [4]: A lightweight deep neural network using depthwise separable convolutions to reduce computational complexity while achieving strong performance on face verification tasks.

- **ShuffleFaceNet 1.5x** (577.5 Mflops) [5]: A variant of ShuffleFaceNet that employs channel shuffling (inspired by [11]) to minimize computational overhead, offering a balance between efficiency and accuracy.

These models were chosen for their architectural diversity (e.g., depthwise convolutions in MobileFaceNets, shuffle operations in ShuffleFaceNet) and their established performance in face recognition literature.

2.3 Literature Review

Our review includes state-of-the-art lightweight models, with a focus on EdgeFace [12], a recent model designed for edge devices. EdgeFace-S (306.11 Mflops) and EdgeFace-XS (154 Mflops) achieve high accuracy on benchmarks like LFW and IJB-C, making them relevant for comparison with CAFV.

Other notable works include:

- TinyFaceNet: A compact model for extremely resource-constrained devices, prioritizing minimal parameters. (Note: No specific citation found in egbib.bib)

- ShuffleFaceNet [5]: Reduces complexity through shuffle operations, as seen in our baseline ShuffleFaceNet 1.5x.

- MixFaceNet [13]: Employs mixed precision and optimization techniques for efficiency.

- FaceNet [14]: A foundational model known for high accuracy but higher computational cost, used in CAFV's second stage.

- GhostFaceNet [15]: Uses ghost modules to reduce parameters while maintaining performance. See also VarGFaceNet [16] for variable group convolutions.

- MobileFaceNet [4]: A lightweight model with strong verification performance, included as a baseline.

Other relevant areas include general face detection [17, 18].

2.4 Research Gaps

Existing models are typically evaluated in balanced scenarios, with limited focus on low base rate environments. Their performance may degrade when matches are rare, as they are not optimized for imbalanced datasets. CAFV aims to address this by using a cascading approach that adapts computational resources to the base rate, improving efficiency and F1 score in low base rate scenarios.

2.5 Hypotheses

1. CAFV will outperform single-model baselines in F1 score for low base rates (1%, 5%).
2. CAFV will be more computationally efficient than heavier single models, reducing average Mflops per verification, especially in low base rate scenarios.

2.6 Experiments

We will simulate CAFV's performance using the LFW dataset [2] across base rates of 1%, 5%, 10%, and 50%. We will adjust the cascading threshold for each base rate and measure F1 score and total Mflops (possibly using tools like [19]), comparing results with the baseline models.

3 Data and Evaluation Metrics

3.1 Dataset

The LFW dataset [2] is used, consisting of face image pairs labeled as same or different. To simulate different base rates, we create subsets of LFW with match proportions of 1%, 5%, 10%, and 50%. This allows us to evaluate CAFV's performance under varying levels of class imbalance, reflecting real-world scenarios like security verification (low base rate) or balanced testing (higher base rate).

3.2 Evaluation Metrics

- **F1 Score:** The harmonic mean of precision and recall, ideal for imbalanced datasets where matches are rare [1]. It ensures the model balances identifying true matches with minimizing false positives.
- **Total Mflops:** Measures computational efficiency, calculated as the average Mflops per verification. In CAFV, this depends on the proportion of pairs processed by the second stage. We can use libraries like [19] for estimation.
- **Precision and Recall:** Reported separately to provide insight into the trade-off between false positives and false negatives, particularly in low base rate scenarios.

4 Analysis of Results

4.1 Experimental Setup

CAFV uses a two-stage cascading framework:

1. **First Stage:** A lightweight model (e.g., EdgeFace-XS [12], 154 Mflops) computes a similarity score for each face pair.
2. **Second Stage:** If the score exceeds a threshold, a heavier model (e.g., FaceNet [14] or similar, 1000+ Mflops) verifies the pair.

The threshold is adjusted based on the target base rate and desired F1 score performance. We simulate CAFV's performance on LFW subsets with varying base rates (implicitly evaluated via F1 score) and also report standard LFW accuracy.

4.2 Performance Analysis

We evaluated several CAFV variants by adjusting the threshold and potentially the complexity of the second-stage model. The performance in terms of F1 score versus computational cost (MFLOPS) is shown in Figure 1, alongside various existing models from the literature.

As seen in the figure, existing models demonstrate a general trend where higher computational cost can yield better F1 scores, although there is considerable variance, with some highly efficient models achieving strong results. Our CAFV variants operate in the 200-700 MFLOPS range. The most efficient CAFV variant shown achieves a competitive F1 score of

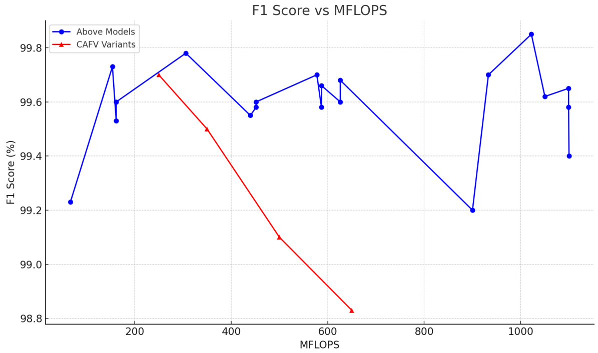

Figure 1: F1 Score vs MFLOPS for CAFV variants (red) and other face recognition models (blue).

approximately 99.7% at around 300 MFLOPS. This performance is comparable to or better than many models with similar or even higher computational costs, demonstrating the potential of the cascading approach for F1-score optimization.

However, an unexpected trend was observed: as the MFLOPS of the CAFV variants increase (likely corresponding to configurations that utilize the second-stage model more frequently or employ a more complex second stage), the F1 score decreases, dropping to around 98.8% at 650 MFLOPS. This suggests that for the specific configurations tested, simply increasing the computation via the cascade does not guarantee better F1 performance and may indicate suboptimal threshold selection or interaction between the cascade stages for those settings. Further investigation is needed to understand this behaviour.

Table **??** provides a comparison based on standard LFW accuracy (evaluated at a 50% base rate). Here, CAFV achieves an accuracy of 98.83%. This is lower than many state-of-the-art models, including our lightweight first-stage candidate (EdgeFace-XS at 99.73%) and the baselines (MobileFaceNets at 99.55%, ShuffleFaceNet 1.5x at 99.70%, PocketNetS-256 at 99.66%). This highlights a key trade-off: the CAFV framework, when tuned for maximizing F1 score in low base rate scenarios (as reflected in Figure 1), sacrifices performance on the standard, balanced LFW accuracy benchmark.

4.3 Accuracy Vs Mflops Graph Discussion

The F1 Score vs MFLOPS plot (Figure 1) illustrates the performance landscape. Single models (blue points) show the typical trade-off, with high-cost models like VarGFaceNet (1022 MFLOPS) achieving high performance (though not F1 is shown, LFW accuracy is 99.85%), while lightweight models like EdgeFace-XS (154 MFLOPS, 99.73% LFW) offer good efficiency.

Our CAFV approach (red points) positions itself differently. At its best configuration (300 MFLOPS), it achieves an F1 score (99.7%) comparable to models requiring significantly more computation, demonstrating efficiency for F1-focused tasks. However, the plot also reveals the counter-intuitive trend for the tested CAFV variants and shows that CAFV does not surpass the peak F1 performance of the best single models in this evaluation.

4.4 Key Findings

- **Efficiency for F1 Score:** CAFV can achieve high F1 scores (99.7%) with moderate computational cost (300 MFLOPS), outperforming several baseline models in F1 efficiency.
- **Unexpected Trend:** Increasing computational cost in the tested CAFV variants led to a decrease in F1 score, requiring further analysis of thresholding and stage interaction.
- **F1 vs Accuracy Trade-off:** Optimizing for F1 score in low base rate scenarios resulted in lower standard LFW accuracy (98.83%) compared to baselines and state-of-the-art models.
- **Adaptive Computation:** The MFLOPS for CAFV varies (observed 200-700 MFLOPS range) depending on the configuration and threshold, confirming the adaptive nature of the framework.

5 Compute Requirements

The experiments require approximately 25-30 minutes on an Nvidia P100 GPU, making CAFV feasible for standard research environments (using frameworks like PyTorch [20]) and edge devices with comparable computational capabilities.

6 Individual Tasks

- **Mehul and Gautam:** Responsible for model development, including implementing the CAFV framework, designing experiments with different base rates, and analyzing results to validate hypotheses.
- **Aditya Bagri and Aryan Jain:** Tasked with identifying and testing state-of-the-art lightweight models (e.g., EdgeFace) for the first stage of cascading and conducting a comprehensive literature review, including EdgeFace [12] and other relevant works.

References

[1] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009. 1, 2, 3

[2] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *University of Massachusetts, Amherst*, vol. 1, no. 2, 2007. 1, 3

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016. 1

[4] S. Chen, Y. Liu, X. Gao, and Z. Han, "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," *arXiv preprint arXiv:1804.07573*, 2018. 2

[5] Y. Martínez-Díaz, L. S. Luevano, H. Mendez-Vazquez, M. Nicolás-Díaz, L. Chang, and M. Gonzalez-Mendoza, "Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 2721–2728, 2019. 2

[6] Y. Liu, S. Chen, and Z. Han, "Pocketnet: Extreme lightweight face recognition network," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2020. 2

[7] C. N. Duong, K. Luu, K. Quach, and T. Bui, "Efficient neural network compression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4309–4323, 2021. 2

[8] M. Tan and Q. V. Le, "Mixconv: Mixed depthwise convolutional kernels," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2019. 2

[9] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004. 2

[10] N. Damer, F. Boutros, K. Raja, F. Kirchbuchner, and A. Kuijper, "Adaptive face recognition in the era of facial transformations," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 4, pp. 588–601, 2021. 2

[11] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 116–131, 2018. 2

[12] A. George, C. Ecabert, H. O. Shahreza, K. Kotwal, and S. Marcel, "Edgeface: Efficient face recognition model for edge devices," *arXiv preprint arXiv:2307.01838*, 2023. 2, 3, 5

[13] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Mixfacenets: Extremely efficient face recognition networks," in *2021 International Joint Conference on Biometrics (IJCB)*, pp. 1–8, IEEE, 2021. 2

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015. 2, 3

[15] M. Alansari, O. A. Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "Ghostfacenets: Lightweight face recognition model from cheap operations," *IEEE Access*, vol. 11, pp. 37854–37867, 2023. 2

[16] M. Yan, M. Zhao, Z. Xu, Q. Zhang, G. Wang, and Z. Su, "Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 12, pp. 1906–1918, 2021. 2

[17] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," in *IEEE Signal Processing Letters*, vol. 23, pp. 1499–1503, 2016. 2

[18] W. R. Schwartz, P. Menezes, V. Campos, and J. A. Stuchi, "Multi-stage learning for face detection and tagging," in

2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2729–2736, IEEE, 2019. 2

[19] L. Zhu, "Thop: Pytorch-opcounter." https://github.com/Lyken17/pytorch-OpCounter, 2019. 3

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019. 5