Robust Face Recognition and Gender Classification under Adverse Conditions using CLIP, DeepFace, and ChromaDB

[1]Soham Chaudhuri,
[2]Sannidhya Das

[1]Jadavpur University, dept.of EE
[2]St. Xavier's College, Kolkata, dept. of Data Science

e-mail: [1] sohamchaudhuri.12.a.38@gmail.com,
[2] sannidhyadas0howrah@gmail.com

Abstract - Real-world face analysis remains a significant challenge, particularly for face recognition tasks under adverse visual conditions such as motion blur, poor lighting, and environmental distortions. In this work, we propose a robust and efficient dual-task pipeline for gender classification and face recognition tailored to such scenarios. For Task A (Gender Classification), we leverage a **pre-trained CLIP image encoder** combined with a lightweight custom classifier, trained specifically on the clean male and female face images provided in the **FACECOM dataset**. This approach enables effective gender prediction without being impacted by environmental degradations. For Task B (Face Recognition), we adopt a **lightweight, training-free solution** using the **DeepFace framework** with ArcFace embeddings to represent identity. We integrate **ChromaDB**, a high-performance vector database, for fast and scalable face retrieval. To address the adverse conditions present in Task B, we apply classical preprocessing techniques including denoising and contrast-limited adaptive histogram equalization (CLAHE) for lighting normalization, ensuring consistent input quality. Evaluation on the FACECOM dataset demonstrates the robustness of our pipeline, with competitive performance achieved in both gender classification and identity recognition tasks. Our system offers a practical, scalable solution for facial analysis in uncontrolled environments, with potential applications in surveillance, forensics, and human-computer interaction.
For further details, visit `(https://github.com/Soham-Chaudhuri/COMSYS-5/tree/main)`

1 Task A: Gender Classification using CLIP

For gender classification, we replace traditional black-box analyzers with a custom, fine-tuned pipeline built on top of the **CLIP (Contrastive Language-Image Pretraining) ViT-B/32 model**. Specifically, the image encoder of CLIP is utilized to extract high-dimensional feature representations from denoised and preprocessed face images. A lightweight linear classifier is trained on top of these frozen image features to perform binary gender prediction.

To enhance robustness, face images undergo aggressive preprocessing, including strong denoising via Non-Local Means filtering and image augmentations such as random flips and rotations. The classifier is trained with class-weighted cross-entropy loss to mitigate dataset imbalance. Our experiments on the **FACECOM dataset** demonstrate that this CLIP-based approach offers improved generalization under degraded conditions compared to generic pretrained gender analyzers, achieving strong **F1-scores and high recall** for both genders.

Set	Metric	Female	Male	Overall
Train	Precision	0.91 (303 samples)	1.00 (1623 samples)	-
	Recall	0.99	0.98	-
	F1-Score	0.95	0.99	-
	Accuracy	-	-	0.98
	Macro Avg. F1	-	-	0.97
	Weighted Avg. F1	-	-	0.98
Validation	Precision	0.74 (79 samples)	0.94 (343 samples)	-
	Recall	0.76	0.94	-
	F1-Score	0.75	0.94	-
	Accuracy	-	-	0.91
	Macro Avg. F1	-	-	0.85
	Weighted Avg. F1	-	-	0.91

Table 1: Gender Classification Results on Train and Validation Sets

2 Task B: Face Recognition with ArcFace and ChromaDB

For identity recognition, we retain the efficient, training-free pipeline based on the **DeepFace** framework and **ChromaDB**. Face embeddings are extracted using DeepFace's **ArcFace** model and stored in ChromaDB, a high-performance vector database supporting scalable similarity search. Label metadata corresponding to each embedding is stored during the registration phase, where the label denotes the identity (i.e., the person's name). During inference, the embedding of the query image is computed and compared against the stored gallery embeddings using *cosine similarity*, enabling fast and robust identity retrieval even under challenging visual distortions such as occlusion or illumination changes. To enhance reliability, a *top-k* retrieval strategy is employed, followed by a *majority voting* scheme on the top retrieved labels to determine the final predicted identity.

Set	Precision	Recall	F1-Score	Accuracy
Training	1.00	0.94	0.97	0.94
Validation	1.00	0.94	0.97	0.94

Table 2: Classification performance for the `recognized` class

3 CONCLUSIONS

This work presented a robust pipeline for gender classification and face recognition under degraded visual conditions. The proposed CLIP-based gender classifier achieved **98% accuracy** with a **macro F1-score of 0.97** on the training set, and **91% accuracy** with a **macro F1-score of 0.85** on the validation set. For face recognition, the DeepFace and ChromaDB system achieved **94% accuracy** and an **F1-score of 0.97** consistently on both the training and validation sets. These results demonstrate the effectiveness of combining pretrained models, custom classifiers, and scalable vector search for reliable facial analysis in challenging real-world scenarios.

References

[1] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pat-

tern recognition (pp. 1701-1708). https://ai.meta.com/research/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/

[2] Y. Chang, C. Jung, P. Ke, H. Song and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization With Dual Gamma Correction," in IEEE Access, vol. 6, pp. 11782-11792, 2018, doi: 10.1109/ACCESS.2018.2797872. https://ieeexplore.ieee.org/abstract/document/8269243

[3] Astawa, I. Nyoman Gede Arya, et al. "Face Images Classification using VGG-CNN." Knowl. Eng. Data Sci. 4.1 (2021): 49-54. https://journal2.um.ac.id/index.php/keds/article/view/19547

[4] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. https://ieeexplore.ieee.org/document/7298682

[5] Deng, Jiankang, and Stefanos Zafeiriou. "Arcface for disguised face recognition." proceedings of the IEEE/CVF international conference on computer vision workshops. 2019. https://ieeexplore.ieee.org/document/9022036

[6] Mosayyebi, Faezeh, Hadi Seyedarabi, and Reza Afrouzian. "Gender recognition in masked facial images using EfficientNet and transfer learning approach." International Journal of Information Technology 16.4 (2024): 2693-2703. https://link.springer.com/article/10.1007/s41870-023-01565-4

[7] Muthukumar, Vidya, et al. "Understanding unequal gender classification accuracy from face images." arXiv preprint arXiv:1812.00099 (2018). https://arxiv.org/abs/1812.00099