Robust Face Recognition and Gender Classification under Adverse Visual Conditions

Github Repository Link `https://github.com/Ayushsaha004/COSMYS_WIZARDS`
Team Name WIZARDS
Team Members Baiduryya Garai, Ayush Saha, Antara Pal

1 Problem Overview

We address the problem of face analysis under degraded visual conditions such as blur, fog, overexposure, rain, and low light. Our solution tackles two critical sub-tasks: Task A: Gender Classification (Binary) Task B: Face Recognition (Multi-Class)

2 Approach and Methodology

2.1 Preprocessing

For Task A, Images were resized to 224×224 and preprocessed using the built-in function for EfficientNetV2B2, which includes model-specific image preprocessing. It ensured compatibility and consistency for effective transfer learning.

For Task B, Each image was normalized individually using min-max normalization i.e $img_{new} = \frac{img - img_{\min}}{img_{\max} - img_{\min}}$, where $img_{\min}$ and $img_{\max}$ are the minimum and maximum pixel values of that image, scaling all pixel values to the $[0, 1]$ range for robustness to lighting and contrast variations.

2.2 Model Architectures

For Task A, we used transfer learning with EfficientNetV2B2 as the base model, removing its classification head and fine-tuning all layers. The output feature map was reduced via Global Average Pooling and reshaped to (1, 1408) as input to a single-layer LSTM with 64 units to capture spatial-temporal dependencies. A Dense layer with softmax activation performed binary gender classification. The proposed architecture of this model is illustrated in Figure 1.

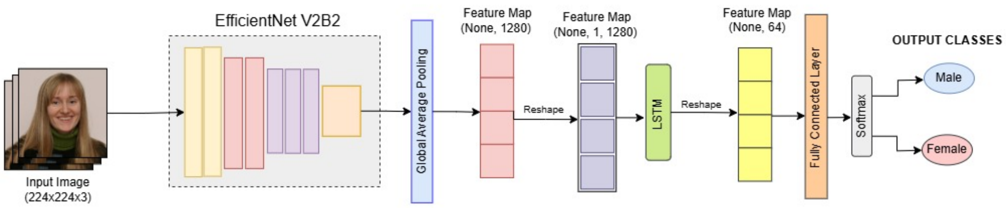

Figure 1: Proposed architecture for Task A using EfficientNetV2B3 and LSTM for gender classification.

For Task B, a Siamese network with two identical convolutional sub-networks based on Resnet50 (without classification heads) extracted fixed-size embeddings from image pairs. The absolute difference of embeddings passed through fully connected layers ending in a sigmoid-activated Dense layer, producing a similarity score indicating identity match. The proposed architecture of this model is shown in Figure 2.

2.3 Training Strategy

The Task A model was trained using the Adam optimizer (learning rate 0.01, epsilon 0.1) with categorical cross-entropy loss. Early stopping based on validation loss was applied over 60 epochs, saving the best checkpoint. Transfer learning leveraged pretrained ImageNet weights for EfficientNetV2B3 to enhance feature extraction.
For Task B, the Siamese network was trained on 52,620 image pairs (balanced with similar and dissimilar pairs) from FACECOM training and validation sets. A custom data loader dynamically prepared batches of size 10, using binary cross-entropy loss and the Adam optimizer. Training comprised 100 batches per epoch with data shuffling and validation monitoring to ensure generalization and prevent overfitting.

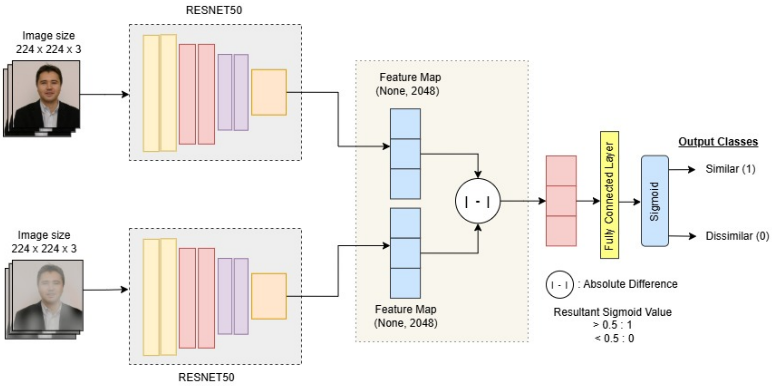

Figure 2: Proposed architecture for Task B using a Siamese network with ResNet50 as the shared feature extractor for face verification.

3 Innovation and Highlight

Task A – Gender Classification

- Combined EfficientNetV2B3 with LSTM to model spatial-temporal features.
- Fine-tuned full EfficientNetV2B3 for domain-specific adaptation.
- Handled class imbalance via architecture and early stopping, not oversampling.

Task B – Face Recognition

- Reformulated task as verification using a Siamese network.
- Applied balanced pair sampling for effective training.
- Compared embeddings via absolute difference for lightweight similarity.

4 Results

For Task A, The model achieved an overall accuracy of 93.6%, with class-wise accuracies of 86.1% for Male and 95.3% for Female, demonstrating strong and balanced performance. Detailed results are presented in Table 1.

Table 1: Performance Metrics and Class-wise Accuracy

Overall Performance				
Dataset	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Train	1.0	1.00	1.00	1.00
Validation	0.936	0.94	0.94	0.94

Class-wise Accuracy		
Class	**Train**	**Val**
Male	1.00	0.861
Female	1.00	0.953

For Task B, using the Siamese-based verification model adapted for identity prediction, we achieved a **Top-1 accuracy of 97.66%** and **macro average f1-score is 92.17%**. The model showed strong generalization with a validation accuracy of 92.17%, as detailed in Table 2.

Table 2: Performance Metrics for Task B – Face Recognition

Dataset	Accuracy	Precision	Recall	F1-Score
Train	99.55%	99.55%	99.55%	99.55%
Validation	92.17%	92.17%	92.17%	92.17%

5 Conclusion and Future Work

We presented robust models for gender classification and face recognition under challenging visual conditions using the FACECOM dataset. Task A achieved 93.6% validation accuracy using an EfficientNetV2B3 + LSTM hybrid, while Task B, reformulated as a Siamese verification task, reached 97.66% Top-1 accuracy.

In future work, we plan to explore advanced augmentation, visual restoration techniques, and metric-learning strategies like triplet loss to further improve performance and generalization.