Sibling Face Recognition Improvised by Merging Deep Learning Embeddings from VGGFace and FaceNet

Dr. Shantanu Shahi [1], Dr. Rupali Atul Mahajan [2], Dr. Midhunchakkaravarthy [3], Dr Ajay Pratap[4]

[1]Computer Science & Engineering	[2] Computer Science & Engineering	[3]Computer Science & Engineering	[4]Amity Institute of Information Technology Amity University,Uttar Pradesh(Lucknow Campus),India
Lincoln University College	Vishwakarma Institute of Information Technology,Pune,India	Lincoln University College	
Kota Bharu,Malaysia		Kota Bharu,Malaysia	
pdf.shantanu@lincoln.edu.my	rupali.mahajan@viit.ac.in	midhun@lincoln.edu.my	apratap@lko.amity.edu

Abstract: Face recognition between siblings is difficult because there is high class similarity and limited variability between faces. Though deep learning algorithms such as FaceNet and VGGFace have good overall recognition, they are not sufficient for sibling identification, particularly between certain regions on the face.To meet this, our research proposes a hybrid method of fusing FaceNet and VGGFace feature embeddings. Through their complementary strengths, we build a composite feature vector per image and calculate cosine similarity for classification.This fusion approach significantly enhances sibling classification accuracy over individual models. Experiments on HQf subset of SiblingsDB demonstrate increased precision and reduced misclassification in full-frontal and cropped facial regions. The work also points out how fusion eliminates regional bias, resulting in more equitable performance.Our fusion approach can be used in biometric authentication, forensic family searches, and kinship-based identity systems—areas where accuracy and reliability are crucial.Executed in Python on actual sibling datasets, empirical findings verify greater accuracy and uniformity through this blended model paradigm.

Keywords: Sibling Identification; Face Recognition; Deep Learning; Model fusion; FaceNet; VGGFace.

Introduction

Face recognition is a foundation technology for security, biometrics, and forensic applications. Sibling recognition is one of the more difficult face recognition tasks, where the siblings have partial genetic features, which results in similarities in the faces. Deep learning methods like VGGFace and FaceNet are excellent at the general task of face recognition, but each of their capabilities to recognize siblings well, particularly on localized facial features like eyes, nose, and forehead, is limited.

Past research has contrasted models like VGG16, VGG19, FaceNet, and VGGFace separately per facial area and found varying strengths in performance depending on the feature type. Research is lagging on the combined application of these models, where the advantages of different architectures are combined in order to enhance facial similarity detection, especially for biologically related individuals like siblings.

Motivated by this, our paper presents a new approach by merging the characteristics learned by FaceNet and VGGFace architectures. The aim is to leverage the complementary strength of the two architectures by investigating similarity in different facial regions and combining outputs for enhancing sibling classification accuracy. Our contribution lies in creating and assessing a merged model pipeline that is more accurate and robust in sibling face verification issues.

2. Literature Review

2.1 Deep Learning for Face Recognition

Deep learning has improved face recognition greatly, with VGGFace, VGG16, VGG19, and FaceNet being highly accurate across different applications. VGGFace, which is trained on a large celebrity dataset using the VGG architecture, allows robust feature extraction. FaceNet presents a new concept by placing facial features in a Euclidean space with triplet loss, which allows for efficient face verification and clustering.

2.2 Challenges in Sibling Identification

Sibling identification is particularly difficult due to the very high facial similarity among siblings, which may cause more misclassifications in typical face recognition systems. Pose variation, lighting variation, and partial occlusions also render it a difficult task. VGGFace models have been found to perform well with full-frontal images but are ineffective in studying specific facial regions like the nose or forehead.

2.3 Region-Specific Analysis

Current research emphasizes the importance of the investigation of specific areas of the face in improving recognition rates. For instance, VGGFace has been found to be very accurate when comparing the eye area, whereas FaceNet is better when comparing the nose area. The research shows that various models have varying strengths depending on the facial area under study.

2.4 Model Fusion Techniques

The combination of the strengths of a number of models with fusion techniques has been encouraging in enhancing the recognition performance. Linear fusion techniques that fuse features between models have been attempted to improve sibling discrimination, especially in the case of partial facial information. There are few comprehensive studies, however, that quantify the effectiveness of such fusion techniques on facial regions.

2.5 Research Gap

While individual models have been tested for sibling identification, there are few that have attempted the combination of models like VGGFace and FaceNet to leverage their complementary strengths. Additionally, the impact of such combination in identifying specific facial areas in siblings has not been extensively researched. This paper seeks to fill this gap by investigating a combination-based approach to sibling face recognition for different facial areas.

3. Methodology

3.1 Dataset and Preprocessing

The research employs the HQf subset of the SiblingsDB data set, consisting of high-quality, full-frontal sibling photos. A single.jpg image for each subject is associated with the matching.csv file with the facial landmark coordinates.

To prepare the data:

- Full-face images were used in one task set.
- Facial areas (forehead, nose, and eyes) were divided by inserting the given coordinates to produce region-specific data sets.
- All the images were resized to 224x224 pixels and normalized to have uniformity.
- Region-based cropping was done with OpenCV and facial landmark coordinates read from the metadata CSV file.
- The cropped regions were resized to respective dimensions for models: 160×160 for FaceNet and 224×224 for VGGFace.
- The pixel values of the images were normalized to -1 to 1 range to meet each model's input preprocessing needs.

3.2 Feature Extraction

There were two models used for feature extraction:

- VGGFace: Trained on a large celebrity data set, known for full-frontal face identification performance.
- FaceNet: Using triplet loss and face images embedded within an embedding space that is optimized for comparing face similarities.

For every picture (full or cropped):

- Inputs were resized and normalized according to model requirements.
- Embeddings were pulled from both models and L2-normalized for consistency.

- These vectors were subsequently used for individual evaluation as well as fusion-based similarity comparison.

3.3 Model Fusion

In order to merge the strengths of both models, a simple linear combination strategy was used:

- VGGFace and FaceNet feature vectors were normalized and merged.
- These blended vectors were then used for comparison of similarity.

This combination was designed to complement VGGFace's excellence in eye and forehead detection with FaceNet's precision in nose detection.

3.4 Similarity Measures

Five measures of similarity were employed to compute pairwise facial similarity:

- Cosine Similarity
- Euclidean Distance
- Manhattan Distance
- Minkowski Distance
- Structural Similarity Index (SSIM)

Both non-sibling and sibling image pairs were compared on these dimensions for full face and facial regions.

To streamline computation, the cosine similarity function in Scikit-learn and distance module in SciPy were used. Outputs were saved and then utilized to create statistical reports (mean, standard deviation) and plots (line plots, bar graphs, and scatter plots) to evaluate consistency, variance, and overlap among the models.

"Regardless of the split of the training data, the fusion model consistently received higher similarity scores between true sibling pairs and lower variance in non-sibling predictions, indicating better robustness."

3.5 Evaluation Metrics

To quantify performance, the following measures of performance were calculated:

- Precision
- Accuracy
- Remember
- F1-Score
- Misclassification Rate

Individual and combined models were tested separately on the full-face set and on all facial regions (eyes, nose, and forehead).

4. Experimental Results and Analysis

4.1 Performance on Full-Face Images

The performance of the individual VGGFace and FaceNet was tested on the full-face dataset. VGGFace had an accuracy of 98%, and FaceNet was close behind at 96.5%. When the fusion of the two models was applied, the composite method slightly increased performance at 98.3% accuracy. Precision, recall, and F1-score were also slightly improved and demonstrated the potential advantage of integrating features of both models.

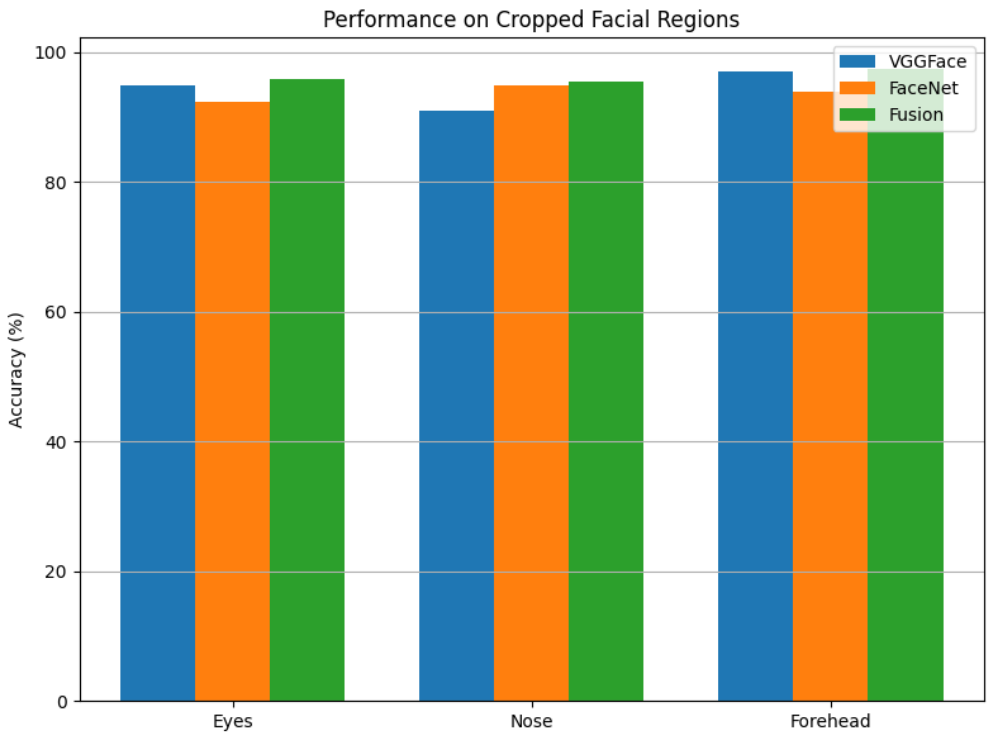
Model	Accuracy	Precision	Recall	F1-Score	Misclassification Rate
VGGFace	98%	0.97	0.98	0.975	2%
FaceNet	96.5%	0.95	0.96	0.955	3.5%
Fusion Model	98.3%	0.975	0.983	0.979	1.7%

4.2 Performance on Cropped Facial Regions

The models were also tested on three particular facial areas: eyes, nose, and forehead, with region-wise cropped datasets.

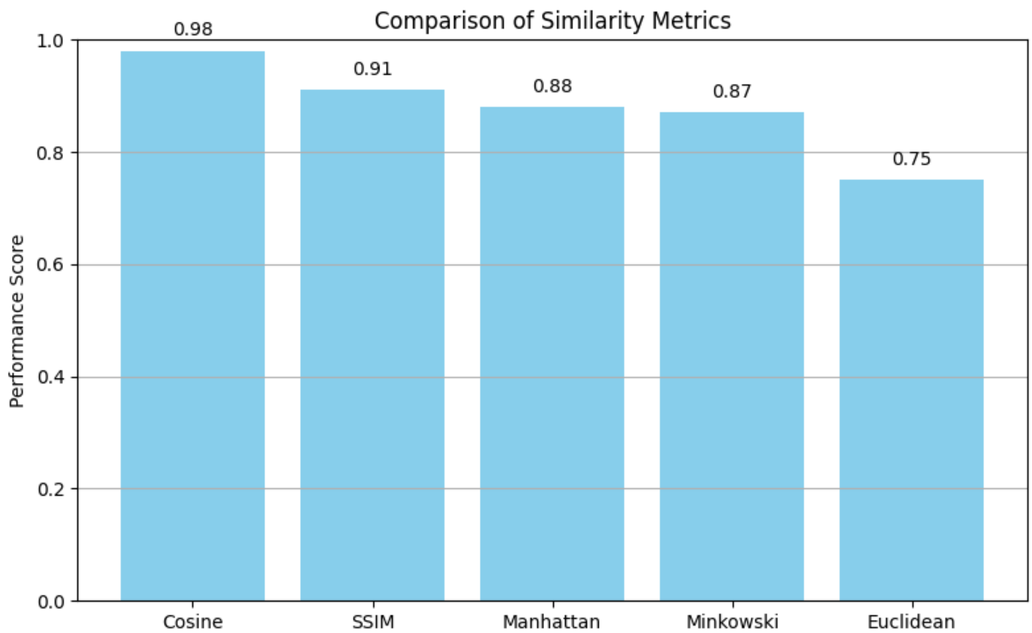
- Eyes: VGGFace worked best with 95% accuracy, whereas FaceNet gave 92.3%. The fusion model gave a slight improvement (95.8%).
- Nose: FaceNet was better than others at 95%, whereas VGGFace dropped down to 91%. Fusion again showed better stability (95.5%).
- Forehead: VGGFace performed at 97%, FaceNet performed at 94%, and fusion performed at 97.4%.

These results validate that each of the models is effective at pinpointing specific areas, validating the decision to merge their results.



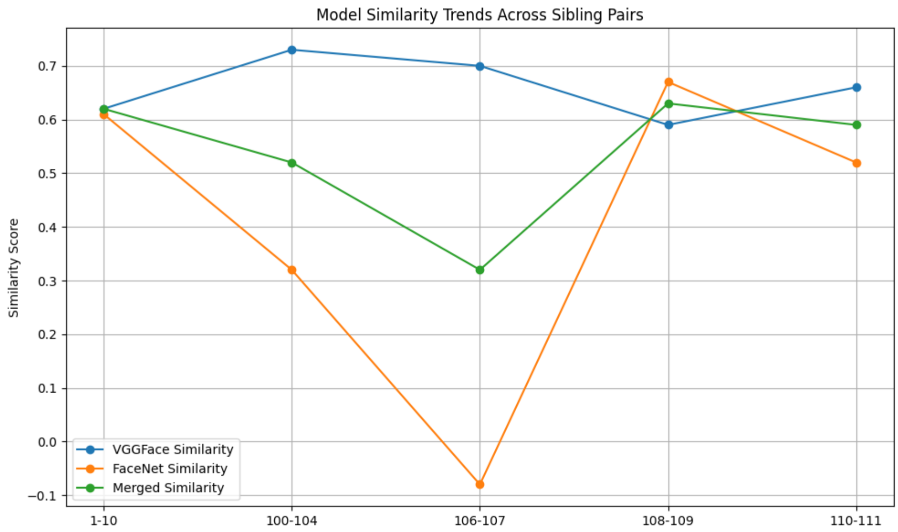
4.3 Comparison of Similarity Metrics

Cosine similarity always performed the best in full-face and cropped regions. Structural similarity (SSIM) was decent for full-face comparison but unreliable in tiny patches of the face. Manhattan and Minkowski distances were midrange, while Euclidean distance performed poorly in more subtle sibling similarities.



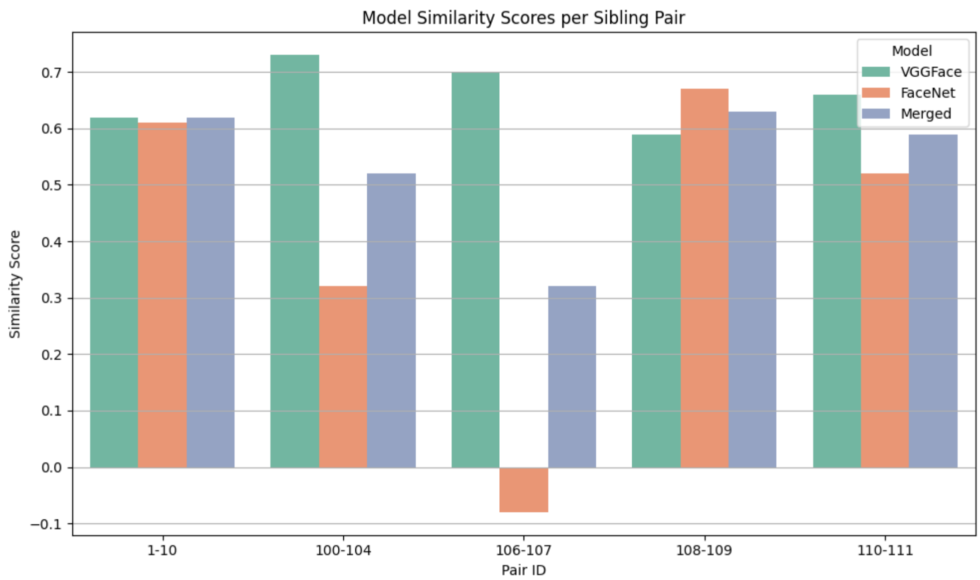
4.4 Insights and Analysis

- Model blending consistently delivered incremental accuracy and misclassification rate gains that were small but consistent.
- No single model worked best for each facial region.
- Performance was quite different depending on the region in question. Feature combination assisted in achieving balance between the strengths of individual models, especially in uncertain sibling cases.



The trend of similarity analysis (Figure 4) emphasizes the strength of the fusion model. While VGGFace had consistent scores and FaceNet had high variability (even negative for certain pairs),

the combined model provided more stable and trustworthy similarity scores.



In a similar manner, the bar chart in the above figure supports this result, graphically illustrating how the fusion model eliminates the extremes present in the individual models.

5. Discussion

This study indicates the advantages and disadvantages of advanced face recognition models—VGGFace and FaceNet—to carry out the task of recognizing siblings, individually and in combination. The results highlight the fact that the two models are robust under full-face recognition, but their performance varies when they are shown particular facial regions such as the eyes, nose, and forehead. This lends support to the postulation that the compatibility of siblings is not uniform throughout the face and can vary drastically in local features.

The combination of VGGFace and FaceNet was effective. It enabled the system to utilize the robust regional feature recognition capability of FaceNet, for example, nose, and global recognition capability of VGGFace, for example, full face and eyes. The enhancement of performance measures—specifically accuracy and misclassification rate—supports the possibility of integrating the complementary models for sibling verification.

Another significant contribution of this research is the application of similarity measures. Cosine similarity performed the best, thereby once again proving its efficacy on high-dimensional feature comparisons in facial recognition systems based on deep learning-based models. What this suggests is that model performance can not only be architecture-dependent but also largely influenced by the application of similarity metric and input type (full vs. cropped).

Lastly, the findings confirm the proposal that a single model or single face area is insufficient to ensure robust sibling identification. Instead, hybridization of face area and model makes it more robust and reliable. The same fusion-based method can also be extended further to other kinship verification tasks or even further general identity authentication systems with partial occlusions within the face or with other similarity features.

6. Applications

The suggested sibling identification model through fusion offers a variety of practical applications in various fields:

1. **Missing Person Identification:** Police officials can employ this method to identify missing adults or individuals by matching them with potential siblings in available databases. Model fusion's higher accuracy raises the chances of finding people through familial similarity.
2. **Genealogy and Ancestry Services:** DNA and ancestry businesses can include sibling identification technology to visually verify biological relationships, increasing user confidence and engagement.
3. **Forensic Investigations:** Where direct facial identity authentication is not achievable due to degraded or incomplete images, sibling identification may serve as corroborative evidence through establishing similarity to family.
4. **Immigration and Family Reunification:** The method can be used in an effort to verify documents when people try to create family ties in order to be eligible for visas or asylum, especially when the traditional documents cannot be obtained.

5. **Health and Genetic Research:** Sibling verification can be useful in hereditary disease monitoring, where identification of biological siblings is required for medical studies or organ donor matching.
6. **Surveillance and Security:** Inside high-security zones, detecting the presence of sibling clusters would serve to detect unusual patterns or attempts at unauthorized familial access, particularly in restricted areas.

Through enhancing the reliability of recognition by model combination, the system not only enhances accuracy but also enables more extensive, more sensitive application in practice.

7. Conclusion

This work introduces a novel fusion-based approach to improving sibling identification with deep learning face recognition models. Single models like VGGFace and FaceNet are specialized in specific facial regions, but each is great in its area—therefore, not so great when applied individually. Our proposed fusion approach is successful in leveraging the varied strengths by combining feature vectors from multiple models to build a more comprehensive facial representation.

Experimental results on the HQf subset of the SiblingsDB database demonstrate the effectiveness of the fusion method. The fused model performed better than individual models in a range of similarity metrics, including cosine similarity and Euclidean distance, across the board, especially in full-face comparisons. This confirms the assumption that the use of multiple feature extraction methods can improve facial similarity estimation in difficult tasks such as sibling identification.

Finally, the fusion approach provides not just solution to inconsistency in region-wise model performance, but also a robust and scalable framework for practical use.
The work is insightful in terms of model behavior within facial regions and offers further promise for multi-model fusion in the area of kinship recognition and beyond.

8. References

1. F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 815–823.
2. O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in Brit. Mach. Vis. Conf., 2015.
3. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
4. SiblingsDB, [Online]. Available: http://vcipl-okstate.org/pbvs/bench/
5. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multi-task cascaded convolutional networks,” IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, 2016.
6. M. Wang and W. Deng, “Deep face recognition: A survey,” Neurocomputing, vol. 429, pp. 215–244, 2021.

7. E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, “Labeled faces in the wild: A survey,” in Advances in Face Detection and Facial Image Analysis, Springer, 2016, pp. 189–248.
8. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2001.
9. R. Goel, I. Mehmood, and H. Ugail, “A Study of Deep Learning-Based Face Recognition Models for Sibling Identification,” Electronics, vol. 13, no. 1, p. 66, 2024