Developing Facial Recognition-Based Reconnaissance to Enhance Stadium Security and Detect Banned Fans at Soccer Stadiums

Timothy Agboada*, Isaac Osunmakinde*, Claude Turner*, Kingsley Nwosu*
**Department of Computer Science*, *Norfolk State University*, Norfolk, VA, USA
Emails: tayagboada@gmail.com, ioosunmakinde@nsu.edu, cturner@nsu.edu, kcnwosu@nsu.edu

***Abstract*—Soccer or football outside of North America remains a major global industry but faces ongoing issues with hooliganism and stadium violence, which often involves banned fans. Despite the severity of these challenges, research on stadium security is limited with most studies utilizing Convolutional Neural Networks (CNNs) for tasks such as attendance and emotion detection rather than crowd security. This research introduces a facial recognition system aimed at identifying banned soccer fans to enhance stadium security. This system leverages Siamese CNNs for accurate facial detection and matching using publicly available data. Data preprocessing and augmentation strategies improved the model resilience, which leads to a detection accuracy of 94.45% for unoccluded faces and 89.85% for occluded faces as it surpasses traditional CNN models by a significant margin. Additional testing addressed demographic bias and robustness to occlusions, which results in enhanced fairness and balanced accuracy across different racial groups. This research offers a reliable and unbiased facial recognition-based reconnaissance solution tailored to soccer stadium security. The findings can guide the development of best practices for the deployment of facial recognition technologies in sports venues by contributing to safer and more secure environments.**

***Index Terms*—Facial Recognition, Siamese Network, Demographic Bias, Occlusion Handling**

I. Introduction

A. Motivation

Hooliganism and violence in soccer stadiums pose serious concerns, affecting fan safety, club reputation, and the sport's image. Incidents involving re-offenders remain prevalent, tarnishing the sport and deterring families from attending matches [1]. Traditional measures like police presence and surveillance cameras have proven inadequate in preventing banned fans from re-entering stadiums. Facial recognition technology offers a promising alternative but faces challenges in accuracy, bias, and privacy [2]. Fig. 1 illustrates the severity of hooliganism, with arrests reaching thousands in English soccer competitions.

Facial recognition systems excel in controlled environments but struggle in real-world settings due to lighting, crowd density, and occlusions like face masks. Biases affecting certain demographics further challenged their fairness, necessitating diverse datasets and advanced algorithms. This research investigated the feasibility of leveraging facial recognition to identify banned fans, addressing technical and ethical challenges.

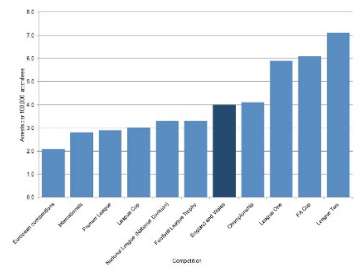

Fig. 1. Arrests at English soccer competitions between 2016 and 2020 adapted from [3]

B. Problem Statement

Hooliganism continues to plague soccer stadiums worldwide, affecting the English soccer hierarchy's image and diminishing the fan experience. The sport faces growing concerns over its family-friendliness due to persistent incidents of violence and misconduct.

C. Research Objectives and Scope

This study aimed to prototype an effective facial recognition system to detect banned fans, improve fairness across demographics, and enhance stadium security. The research replicated typical stadium entry conditions, incorporating diverse lighting and occlusion scenarios, while respecting privacy regulations by using synthetic banned face representations.

D. Research Questions and Contributions

The primary research question asked how a facial recognition-based reconnaissance system could be prototyped to detect banned fans across demographics effectively. Key contributions include:

- A Siamese network model that achieved 94.45% accuracy for unoccluded faces and 89.85% for occluded faces, outperforming traditional CNNs.
- Bias mitigation strategies that improved fairness, with African face accuracy increasing from 75.89% to 82.28%.
- Demonstration of robust handling of occlusions using synthetic augmentation and feature fusion, with minimal accuracy drops compared to traditional methods.

II. Background

A. Related Literature Reviews

This review examined the growing interest in facial recognition technology across various fields and its challenges in implementation. Although limited research exists on detecting banned soccer fans, advancements in facial recognition for attendance systems highlight its potential. However, technical issues such as variations in lighting, facial expressions, and biases hinder accuracy. Table I summarizes pertinent research, key findings, and limitations.

TABLE I
SUMMARIZED RELATED WORKS

Literature	Study	Methodology	Key Results	Limitations
[4]	Video Face Recognition Using Siamese Networks	Evaluated Deep SiamSRC network on Chokepoint and COX-S2V datasets	Performance improved with face synthesis	Domain distribution differences
[5]	Facial emotion recognition based on LDA and Landmark Detection	Comparative study of LDA and Facial Landmark Detection	Landmark detection performed 10.6% better than LDA	Limited dataset variety
[6]	Facial image deviation estimation	Facial image deviation and selection algorithm (FIDE-ISA)	Achieved 100% accuracy with fewer input images	Lacked comparisons with other models
[7]	Multi-pose face recognition	Convolutional Siamese Architecture	Outperformed competing algorithms	High computational cost
[8]	Occlusion handling in face recognition	Pairwise Differential Siamese Network	Superior performance under occlusions	Limited occlusion types (e.g., sunglasses, scarves)

B. Research Gaps

Despite progress, research gaps remain, such as the limited use of facial recognition for soccer stadium security, lack of demographic-aware training for fairness, and minimal handling of occlusions in real-world scenarios.

C. Bias and Occlusion Solutions

Addressing bias requires representative datasets and techniques like stratified sampling, which was chosen for this study to ensure a fair representation of demographics [9] [10]. To handle occlusions, data augmentation proved most effective by simulating varied occlusion scenarios, improving the model's robustness in real-world conditions [11].

D. Deep Learning Techniques

Among the deep learning techniques considered, Siamese networks were selected for their one-shot learning capabilities and ability to measure similarity between image pairs, making them ideal for scenarios with limited data [12] (Table II). While CNNs and FaceNet are widely used, Siamese networks demonstrated superior performance in distinguishing banned fans, particularly under challenging conditions like occlusions [13].

TABLE II
STRENGTHS AND WEAKNESSES OF SIAMESE NETWORKS

Study	Strengths	Weaknesses
[12]	Effective for one-shot learning with limited data; suitable for facial ID and verification	Requires careful similarity tuning; computationally intensive
[11]	Robust in few-shot learning; handles lighting and pose variations well	Struggles with diverse datasets; complexity increases with dataset size

III. Proposed Facial Recognition Reconnaissance

The model focuses on providing two input images and ascertaining their similarity based on a similarity score, as seen in Fig. 2.

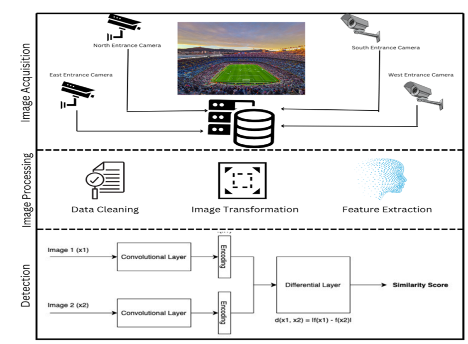

Fig. 2. Siamese Twin Convolutional Neural Network Schema

A. Image Processing

1) Data Cleaning: The image enhancement techniques applied are: Bilateral Filtering (diameter 9 pixels, sigma color 75, sigma space 75); Image Sharpening (3x3 kernel filter [[0, -1, 0], [-1, 5, -1], [0, -1, 0]]); Contrast Enhancement (CLAHE, clip limit 3.0, tile grid 8x8).

2) Image Transformation: 200x200 pixels transformed to 128x128 pixels by: Auto-Cropping (central region); Resizing (bilinear interpolation); Normalization ([0, 1] range).

3) Feature Extraction: Foreground extraction: $F(x,y) = |I(x,y,t) - I(x,y,t-1)|$, where I (x, y, t) is pixel intensity at (x, y) at time t.

4) Data Augmentation: Random Rotations (-15° to +15°); Horizontal Flipping (50% probability); Zooming (60% to 120% range).

B. Detection

1) Convolutional Layers: EfficientNetB0 architecture [14] with ReLU activation: Conv Layer 1 (32 filters, 3x3, stride 1); Conv Layer 2 (64 filters, 3x3, stride 1); Conv Layer 3 (128 filters, 3x3, stride 1).

2) Pooling Layers.: Global Average Pooling Layer; Intermediate Pooling Layers.

The L1Dist layer computes the distance between feature representations from the Embedding layer.

3) Siamese Network Architecture.: Table III provides an overview of the Siamese network architecture:

TABLE III
SIAMESE NETWORK ARCHITECTURE

Layers (type)	Input Shape	Output Shape
Input_img (InputLayer)	(None, 128, 128, 3)	(None, 128, 128, 3)
Validation_img (InputLayer)	(None, 128, 128, 3)	(None, 128, 128, 3)
Embedding (Functional)	(None, 128, 128, 3)	(None, 512)
L1_Dist (Lambda)	(None, 512)	(None, 512)
Dense (Dense)	(None, 512)	(None, 1)

The Siamese network consists of: Input Layers accepting 128x128x3 images; Embedding Layer using EfficientNetB0 for feature extraction, producing 512-dimensional vectors; L1 Distance Layer computing Manhattan distance between feature vectors; Dense Layer for final classification, outputting a similarity score (1 for same person, 0 for different). L1 distance is preferred for its robustness to outliers and computational efficiency [15]. This architecture effectively distinguishes subtle facial differences.

C. Occlusions and Biasness

The model addresses two main challenges: occlusions and bias of response.

1) Occlusions: Data Augmentation is used by introducing synthetic occlusions to simulate real-world scenarios, including random face coverings and noise, enhancing robustness (Fig. 3).

Fig. 3. Black Rectangles Occlusion Data Augmentation

Feature Fusion: Siamese networks use image pairs (one with occlusions, one without) to improve model adaptability. Augmentation techniques like partial face covering or adding noise dots are applied to one image. This trains the model to handle various occlusion levels, ensuring performance when facial features are partially obscured.

2) Biasness: In establishing Race bias awareness, a balanced dataset with diverse skin tones (Fig. 4) mitigates bias. Data augmentation, balanced class sampling, cross-validation, and post-training fairness tests ensure consistent performance across demographics.

Regularization: Dropout rate 0.5 and weight decay factor 0.001 prevent overfitting and improve generalization [16].

D. Training and Evaluation Metrics

Data split: 80% training, 20% testing. Model compiled with binary cross-entropy loss, accuracy metric, and Adam optimizer (learning rate 0.0001). Trained for 13 epochs, with validation monitoring. Post-training metrics: accuracy, precision, recall, F1-score. Additional tests ensured fairness across demographic groups [17].

Fig. 4. Representation of varied skin tones the system receives as input

IV. EXPERIMENTAL EVALUATION AND RESULTS

A. Evaluation Setup and Data Description

The Siamese CNN model was trained on 12,000 images from the UTKFace dataset, resized to 128x128 dimensions. Images were paired to create positive and negative samples, ensuring diversity in age, race, and gender. The dataset curation focused on balanced pair selection, regularization, and adversarial training to address challenges with deranged datasets. Data augmentation introduced occlusions to simulate real-world scenarios. The dataset was divided using stratified sampling, considering age (newborns to 100+), race (Black, Caucasian, Asian, Indian, Others), and gender (male, female). This approach ensured proportionate representation across sets, particularly for age groups, and allowed robust testing of the model's generalization abilities while maintaining demographic balance.

The Siamese CNN model was implemented using Python3 with TensorFlow, Keras, NumPy, and OpenCV. TensorFlow and Keras provided support for Siamese networks, NumPy handled large image arrays, and OpenCV facilitated image pre-processing tasks like resizing and augmentations.

B. Descriptive Analytics from Data

The dataset analysis revealed imbalances across race, age, and gender. Fig. 5 shows race distribution, with underrepresentation in the "Other" group. Fig. 6 illustrates skewed age distribution, particularly among younger females. Fig. 7 and 8 preview images and augmentations, respectively.

Stratified sampling ensured demographic balance across training (80%), validation (10%), and test (10%) sets. Race distribution (40% Caucasian, 20% Black, 15% Asian, 15% Indian, 10% Other) was preserved across subsets. Gender and age groups were proportionally represented to minimize potential biases during training and evaluation.

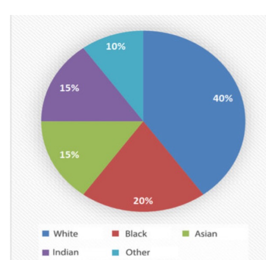

Fig. 5. Distribution of facial images across different demographic groups

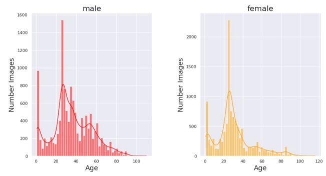

Fig. 6. Histogram depicting the age distribution of individuals by gender

Fig. 7. Sample dataset images, highlighting variations in pose, expression, and illumination

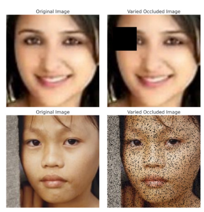

Fig. 8. Examples of augmented images with synthetic occlusions

C. Experiment I: Evaluating Model Performance on Unoccluded Faces

The first experiment for the base case, using unoccluded faces, achieved 94.45% accuracy, indicating the model's strong ability to distinguish between banned and legitimate fans under ideal conditions. This performance, shown in Fig. 9 and detailed in Table IV of the four performance metrics, represents a 6% accuracy improvement over a traditional CNN model [7] that has 88% accuracy.

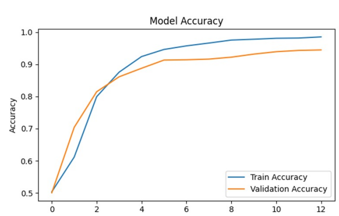

Fig. 9. Graph depicting experiment I model accuracy

TABLE IV
EXPERIMENT I PERFORMANCE METRICS

Metric	Result (%)
Precision	96.15
Recall	92.50
F1-Score	94.29
Accuracy	94.45

D. Experiment II: Evaluating Model Performance on Occluded Faces

Random pixel occlusions simulated real-world scenarios. The model performance decreased compared to Experiment I (Table V, Fig. 10). Multiple randomized occlusions had the most significant impact on model fidelity (Table VI). The accuracy decreased by 4.6% with occluded faces, indicating reduced classification ability for banned soccer fans using multiple occlusions. Performance dropped most when both compared images had occlusions, due to fewer visible features for comparison.

TABLE V
EXPERIMENT II AGGREGATED PERFORMANCE

Metric	Occluded Result (%)	Unoccluded Result (%)
Precision	89.40	96.15
Recall	89.02	92.50
F1-Score	89.21	94.29
Accuracy	89.85	94.45

TABLE VI
EFFECTS OF INDIVIDUAL OCCLUSIONS ON MODEL PERFORMANCE

Occlusion Type	Accuracy (%)	Precision (%)
No Occlusion (Baseline)	94.45	96.15
Partial Eye Occlusion	88.70	86.50
Full Eye Occlusion	82.55	82.34
Mouth Occlusion	79.35	81.39
Forehead Occlusion	73.73	80.00
Both Eyes Occluded	71.13	75.70
Random Multi Pixel Occlusion	71.10	70.91

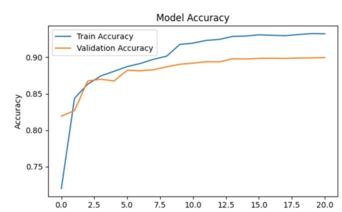

Fig. 10. Graph depicting experiment II aggregated model accuracy

Table VI illustrates the impact of various occlusions on model performance, with larger occlusions significantly reducing accuracy. The experiment focused on maintaining identification accuracy across different races and demographics and evaluating performance before and after applying bias mitigation strategies.

E. Experiment III: Mitigating Demographic Bias in Facial Recognition

Initial Racial Faces in the Wild (RFW) dataset testing revealed performance variations across racial categories (Table VII). The model's accuracy ranged from 88.75% for Caucasian faces to 75.89% for African faces, indicating performance imbalance. The African set's lower accuracy may be due to a smaller age gap and dataset representation.

After implementing bias mitigation strategies, overall model accuracy improved to 89.72% (Fig. 11). Table VIII shows the four performance metrics post-mitigation. The RFW dataset investigated individual class performance, while the LFW dataset provided aggregated results for varied groups.

TABLE VII
RFW RACE ACCURACY BEFORE MITIGATION

Race	Accuracy (%)
Caucasian	88.75
Indian	83.41
Asian	79.62
African	75.89

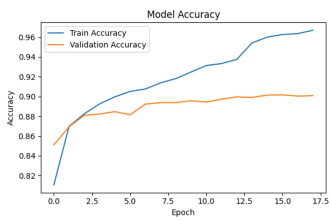

Fig. 11. Graph depicting experiment III model accuracy

TABLE VIII
EXPERIMENT III PERFORMANCE METRICS

Metric	Result (%)
Precision	91.08
Recall	86.95
F1-Score	88.97
Accuracy	89.72

Bias mitigation strategies significantly improved fairness across demographic groups, as demonstrated by retesting the RFW dataset (Table IX). The accuracy gap between racial categories decreased, with the African category showing the most substantial improvement.

TABLE IX
RFW RACE ACCURACY AFTER MITIGATION

Race	Accuracy (%)
Caucasian	89.90
Indian	87.20
Asian	84.87
African	82.28

Cross Validation: K-fold cross-validation ensured model robustness, with metrics averaged across folds to assess generalization and fairness.

F. Performance Evaluations

The confusion matrix (Fig. 12) shows strong overall performance: 962 correct negative and 893 correct positive classifications, with 38 false positives and 107 false negatives. These misclassifications indicate areas for future improvement.

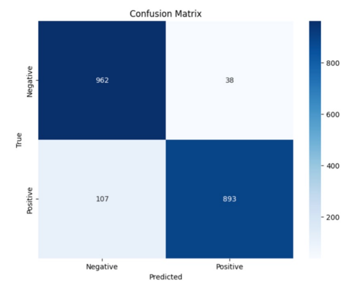

Fig. 12. Confusion Matrix of model performance without occlusions

Fig. 13 shows the confusion matrix for model performance with occlusions, indicating reduced prediction accuracy compared to the non-occluded scenario.

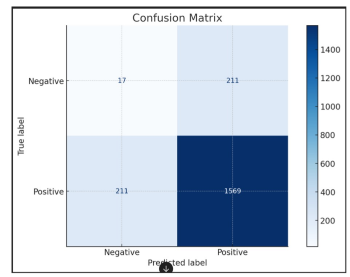

Fig. 13. Confusion Matrix of model performance with occlusions

1) Error Analysis: False positives and negatives are attributed to dataset biases, imbalanced demographic representation, and challenges in distinguishing facial features under varying image quality and lighting conditions.

2) Potential Improvements: Future work could explore increasing training data diversity, focusing on underrepresented groups, implementing advanced data augmentation for lighting variations, and experimenting with deeper architectures or hyperparameter fine-tuning.

3) Generalizability: The model's adaptability was tested on the RFW dataset, demonstrating its ability to generalize to varied inputs and real-world scenarios. Results are shown in Table IX, highlighting the model's performance on this alternative dataset.

G. Comparison to Existing Research

Table X looks at laying out the accuracies of the existing research, as compared to the outcome of this paper. These results indicate that there is a significant improvement made with better resilience in the presence of occlusion, which can help in having better facial reconnaissance.

Additionally, this approach showed a higher resilience as compared to an LLE-CNN approach [18], as the drop in precision without occlusions and with occlusions are 6.75% and 15.6% respectively (smaller is better).

V. ETHICAL AND PRACTICAL DEPLOYMENT IN REAL-TIME

Deploying this framework in stadiums presents challenges such as processing speed, scalability, and robustness in dy-

TABLE X
COMPARISON OF FACIAL RECOGNITION ACCURACY

Article	Accuracy (%)	
	Without Occlusions	**With Occlusions**
Proposed Approach	94.45	89.85
[15]	92.68	Not specified
[19]	93.20	81.30
[8]	91.93	Not specified
[20]	91.40	73.80
[7]	88.45	76.38

namic environments. Solutions include leveraging optimized hardware and parallel processing to enhance real-time performance. Privacy concerns are mitigated by encrypting facial data and adhering to stringent data handling policies, ensuring compliance with legal standards. Ethically, risks of misuse can be addressed by limiting access to authorized personnel and implementing audit mechanisms to ensure accountability in system deployment.

VI. CONCLUSION

This research highlights the potential of a Siamese network-based facial recognition system to combat hooliganism by enhancing stadium security. The proposed model demonstrates superior accuracy and resilience, addressing challenges such as demographic bias and occlusions. Experimental results show its effectiveness in accurately identifying banned individuals, outperforming traditional methods.

While challenges persist, this study underscores the transformative role of facial recognition technology in improving stadium security, reducing violence, and enhancing the fan experience. The insights gained can help restore soccer's family-friendly image, boosting attendance and sponsorships, and paving the way for safer, more enjoyable environments.

Summarized key takeaways from this research are as follows:

- It is established that our model performs well on un-occluded faces, which lays the groundwork for further evaluations on more challenging datasets with occlusions.
- Experiment II highlights the model's reduced results but still notable performance on occluded faces. Further improvements, such as specialized feature extraction for occluded regions could enhance its robustness in real-world scenarios.
- Experiment III demonstrates that while demographic bias was present in the initial model, the application of mitigation strategies significantly reduced performance disparities across demographic groups. Future work could explore additional bias mitigation techniques, such as transfer learning.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the resources made available by Norfolk State University, Virginia, USA.

REFERENCES

[1] R. T. Jewell, R. Simmons, and S. Szymanski, "Bad for business? The effects of hooliganism on English professional football clubs," *Journal of Sports Economics*, vol. 15, no. 5, pp. 429–450, 2014.

[2] B. Constandt, M. M. Parent, and A. Willem, "Does it really matter? A study on soccer fans' perceptions of ethical leadership and their role as 'stakeowners'," *Sport Management Review*, vol. 23, no. 3, pp. 374–386, 2020.

[3] Home Office, "Football-related arrests and banning order statistics, England and Wales: 2016 to 2017 season," GOV.UK, 2017.

[4] G. C. Lee, Z. Y. Li, and T. W. Li, "Ensemble algorithm of convolution neural networks for enhancing facial expression recognition," in *2022 IEEE 5th International Conference on Knowledge Innovation and Invention (ICKII)*, 2022, pp. 111–115.

[5] L. Sun, J. Dai, and X. Shen, "Facial emotion recognition based on LDA and facial landmark detection," in *2021 2nd International Conference on Artificial Intelligence and Education (ICAIE)*, 2021, pp. 64–67.

[6] H. O. Siyaka, O. Owolabi, and I. H. Bisallah, "A new facial image deviation estimation and image selection algorithm (Fide-Isa) for facial image recognition systems: The mathematical models," in *2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)*, 2021, pp. 1–7.

[7] Y. Luo, J. Wu, Z. Zhang, H. Zhao, and Z. Shu, "Design of facial expression recognition algorithm based on CNN model," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, 2023, pp. 580–583.

[8] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential Siamese network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 773–782.

[9] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[10] H. Wu, Z. Xu, J. Zhang, W. Yan, and X. Ma, "Face recognition based on convolution Siamese networks," in *2017 10th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–5.

[11] Z. He et al., "The improved Siamese network in face recognition," in *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, 2019, pp. 443–446.

[12] F. Mokhayeri and E. Granger, "Video face recognition using Siamese networks with block-sparsity matching," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 2, pp. 133–144, 2019.

[13] W. Cui, W. Zhan, J. Yu, C. Sun, and Y. Zhang, "Face recognition via convolutional neural networks and Siamese neural networks," in *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, 2019, pp. 746–750.

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, 2019.

[15] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, "Racial faces in the wild: Reducing racial bias by information maximization adaptation network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 692–702.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[17] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.

[18] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with lle-cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2682–2690.

[19] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2020.

[20] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002–1014, 2018.