Diffusion-Based Synthetic Face Balancing for Fairness in Face Recognition Benchmarks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: To what extent does balancing real and synthetic face images using diffusion models improve fairness metrics across demographic subgroups in deep face recognition benchmarks. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: The Impact of Balancing Real and Synthetic Data on Accuracy and Fairness in Face Recognition. Research question: To what extent does balancing real and synthetic face images using diffusion models improve fairness metrics across demographic subgroups in deep face recognition benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.8/10.

3 Results

14 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 2.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Models trained exclusively on authentic data subsets of CASIA-WebFace without data augmentation showed approximately 15%	×	0.14
Among models trained solely on synthetic images without data augmentation, the model trained on the DCbal subset achieve	×	0.07
The FR model trained on the DCbal synthetic subset outperformed the model trained on the IDFbal subset by an average of	×	0.03
The FR model trained on the DCbal synthetic subset outperformed the model trained on the GCbal subset by 20.16% in avera	×	0.03
The FR model trained on the GCbal subset exhibited a 30.66% reduction in accuracy on the AgeDB-30 benchmark compared to	×	0.04
The FR model trained on the GCbal subset exhibited a 19.39% decrease in accuracy on the CA-LFW benchmark compared to the	×	0.02
ExFaceGAN generated a dataset containing 6,218 Caucasian IDs, 1,973 Indian IDs, 1,668 Asian IDs, and 141 African IDs.	×	0.01
DCFace generated a dataset containing 8,290 Caucasian IDs, 887 Indian IDs, 571 Asian IDs, and 252 African IDs.	×	0.01
IDiff-Face generated a dataset containing 7,464 Caucasian IDs, 1,090 Indian IDs, 915 Asian IDs, and 580 African IDs.	×	0.02
When trained with data augmentation, the model trained on CASIA-WebFace (10K identities) achieved an average verificatio	×	0.04
When trained with data augmentation, combining BUPT and DCbal data (10K total identities) resulted in an average verific	×	0.04

References

- `http://arxiv.org/abs/2402.01472v1`
- `http://arxiv.org/abs/2504.07392v6`
- `http://arxiv.org/abs/2409.02867v1`