Synthetic-to-Authentic Data Mixing for Face Recognition: Impact of Backbone Depth and Loss Functions

Mohamed Mehdi Ayeche
ENSI, HANALAB Lab.
University of Manouba
Manouba, Tunisia
mayeche@yahoo.com

Marouane Ben Haj Ayech
Polytechnique Int. Privée
HANALAB Lab., Tunis, Tunisia
m.ayech@pi.tn

Lotfi Chaouech
ENSIT, LISI Lab.
University of Tunis
Tunis, Tunisia
chaouech.lotfi@yahoo.fr

Ridha Ghayoula
Faculty of Engineering
University of Moncton
Moncton, Canada
ridha.ghayoula@umoncton.ca

Jaouhar Fattahi
Dept. of Comp. Sci. & Soft. Eng.
Laval University
Quebec, Canada
jaouhar.fattahi@ulaval.ca

Lassaad Latrach
ENSI, HANALAB Lab.
University of Manouba
Manouba, Tunisia
lassaad.latrach@ensi-uma.tn

***Abstract*— Data privacy regulations have restricted access to authentic face datasets, making synthetic alternatives increasingly necessary for face recognition training. Yet their joint effect with backbone depth and loss function on verification accuracy remains poorly understood. This paper presents a three-dimensional study using unbalanced subsets of CASIA-WebFace (authentic) and DCFace (synthetic), without demographic correction. The synthetic-to-authentic ratio is varied from 0 to 30 identities, across three backbone depths (ResNet34, ResNet50, ResNet100) and two loss functions (CosFace, ArcFace), evaluated on six standard face verification benchmarks. Three results are reported: (i) synthetic and authentic unbalanced data produce equivalent verification accuracy at all backbone depths, while accuracy decreases with backbone depth under low-data conditions; (ii) the optimal synthetic proportion varies with backbone depth, shifting from 5 identities for ResNet34 and ResNet50 to 15 for ResNet100; (iii) CosFace outperforms ArcFace on synthetic-only data at all depths, while ArcFace outperforms CosFace on combined data for ResNet100, achieving the highest accuracy across all configurations. These results show that backbone depth, mixing ratio, and loss function interact and should not be optimized independently.**

Index Terms—Face recognition, synthetic data, data mixing, backbone depth, margin-based loss functions.

I. INTRODUCTION

Data protection regulations such as GDPR [1] have restricted access to web-crawled face datasets previously used to train recognition systems [2], [3], [4], [5]. Diffusion-based synthetic datasets [6], [7] have been proposed as privacy-friendly alternatives, but their integration into training pipelines raises unresolved questions about mixing ratios, backbone selection, and loss function choice [8]. Both authentic and synthetic collections share a persistent Caucasian overrepresentation [9], [10], and recent work suggests that correcting this imbalance does not consistently improve verification accuracy [11].

Studies on authentic-synthetic data mixing [11], [12] have treated these factors in isolation. Atzori *et al.* [11] evaluated mixing ratios and loss functions on a single backbone (ResNet50) using demographically balanced data, leaving three dimensions uncharacterized: the effect of backbone depth on the optimal mixing ratio, the loss function behavior under unbalanced combined training, and the joint interaction of all three factors.

This paper addresses this gap through a three-dimensional study that jointly examines synthetic data proportion, backbone depth, and loss function on face verification accuracy under unbalanced training conditions. Using subsets of CASIA-WebFace [2] (authentic, 63.4% Caucasian) and DCFace [6] (synthetic, 82.9% Caucasian), we vary the synthetic-to-authentic ratio from 0 to 30 identities, compare three backbone depths (ResNet34, ResNet50, ResNet100) [13], and evaluate CosFace [4] and ArcFace [5] across six standard verification benchmarks.

The contributions of this paper are:

- We show that synthetic and authentic unbalanced data produce equivalent verification accuracy across all three backbone depths, with differences below 1% in all configurations, while average accuracy decreases consistently with backbone depth under the low-data conditions of this study.
- We identify a backbone-dependent optimal mixing ratio: the synthetic complement that maximizes accuracy shifts from 5 identities for ResNet34 and ResNet50 to 15 for ResNet100, a capacity-driven threshold not previously reported.
- We uncover a three-way interaction between loss function, data source, and backbone depth: CosFace outperforms ArcFace on synthetic-only data at all depths, while ArcFace outperforms CosFace on combined data exclusively for ResNet100, achieving the highest accuracy across all configurations (60.54%).

II. RELATED WORK

A. Authentic and Synthetic Data for Face Recognition

Legal and ethical constraints on biometric data collection [1] have driven the transition from web-crawled datasets such as CASIA-WebFace [2] and VGGFace2 [3] toward generative alternatives [8]. GAN-based methods [14], [15], [16] and 3D rendering [12] preceded diffusion-based frameworks including IDiff-Face [7] and DCFace [6], which produce competitive accuracy without real subject data. DCFace is used here because its dual-condition diffusion process is trained on CASIA-WebFace, making it a direct unbalanced synthetic counterpart to the authentic source. Atzori *et al.* [11] showed that authentic-synthetic combinations approach full-authentic accuracy and that demographic balancing yields no consistent gain, motivating the unbalanced design adopted here.

B. Margin-Based Loss Functions

Margin-based losses enforce inter-class separation on the unit hypersphere: CosFace [4] applies a fixed penalty to the target logit in the cosine domain, ArcFace [5] introduces the margin in the angular domain, and AdaFace [17] conditions it on per-sample feature norms. These have been evaluated mainly on large-scale authentic data with fixed backbones; Atzori *et al.* [11] used CosFace exclusively, leaving alternative margin geometries under synthetic or mixed training uncharacterized.

C. Backbone Depth and Data Combination

Residual networks [13] span a well-characterized depth spectrum (ResNet34/50/100). On large-scale data, deeper variants generally achieve higher accuracy [5], [17], but this advantage depends on sufficient data to populate the larger parameter space; in low-data regimes deeper architectures may underperform shallower ones, which becomes relevant under privacy constraints. On the data side, DigiFace-1M [12] showed that a small authentic complement to synthetic sets yields accuracy gains, indicating source complementarity. Atzori *et al.* [11] remain the most directly related study, evaluating mixing ratios on balanced subsets with a single backbone and loss. No prior work has jointly examined synthetic proportion, backbone depth, and loss geometry under unbalanced, privacy-constrained conditions; this paper provides that analysis.

III. THREE-DIMENSIONAL STUDY DESIGN

This section presents the experimental framework designed to jointly investigate the effect of synthetic data proportion, backbone depth, and loss function on face verification accuracy under unbalanced and privacy-constrained training conditions. Fig. 1 provides a visual overview. Both datasets exhibit a pronounced Caucasian overrepresentation (Table II), preserved throughout all experiments since this study targets unbalanced data interaction rather than demographic correction.

TABLE I
POSITIONING WITH RESPECT TO ATZORI *et al.* [11]

Aspect	Atzori *et al.*	This work
Backbone	ResNet50 only	R34 / R50 / R100
Loss	CosFace only	CosFace + ArcFace
Auth. Data	BUPT + CASIA	CASIA only
Synth. Data	GC + DC + IDF	DCFace only
Ratio	Fixed	Variable (0–30)
Scale	Up to 5K IDs	Up to 40 IDs
Dimensions	2 (data, loss)	3 (data, backbone, loss)

A. Research Questions

Three research questions structure this study. Backbone depth (ResNet34, ResNet50, ResNet100) is not an independent research question but a shared experimental dimension that applies to all three RQs, as shown in Fig. 1.

RQ1 – Data Source: Can unbalanced synthetic data (DCFace) match or exceed the verification accuracy of authentic data (CASIA-WebFace) when trained on the same number of identities under identical conditions, across all three backbone depths?

RQ2 – Mixing Ratio: Does combining unbalanced authentic and synthetic data improve verification accuracy over single-source training, and does the optimal synthetic-to-authentic ratio vary with backbone depth? The number of authentic identities is fixed at 10, while synthetic identities are varied from 0 to 30 in steps of 5, across the three backbones.

RQ3 – Loss Function: Does the choice of loss function (CosFace vs. ArcFace) interact with data source and backbone depth? Both loss functions are evaluated across authentic-only, synthetic-only, and optimally combined configurations for each backbone depth.

B. Positioning with Respect to Prior Work

Table I contrasts this study with the most closely related work [11], introducing three dimensions not jointly studied there: backbone depth as an explicit experimental variable, unbalanced data combinations, and systematic variation of the synthetic-to-authentic ratio across multiple backbones and loss functions.

IV. DATASETS AND EXPERIMENTAL SETUP

A. Datasets

Two face datasets are used, both retained in their original unbalanced demographic state. CASIA-WebFace [2] contains approximately 0.5 million face images from 10,000 subjects collected from the web; its acquisition without subject consent raises compliance issues with current data protection legislation [1], motivating the use of synthetic alternatives. DCFace [6] generates synthetic identities through a dual-condition diffusion process conditioned on identity and style images, producing realistic face images without relying on real subject data. Both sources share a pronounced Caucasian overrepresentation [11] preserved throughout all experiments. Experimental subsets of 10 identities per source are summarized in Table II.

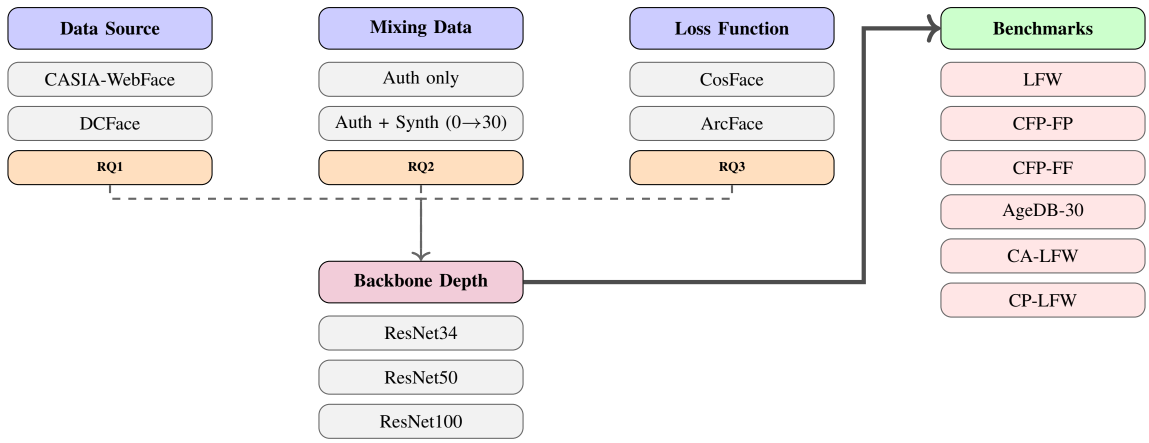

Fig. 1. Overview of the three-dimensional experimental framework. Data sources are used separately (RQ1) or mixed in varying synthetic proportions (RQ2), evaluated across two loss functions (RQ3). Backbone depth acts as a shared experimental dimension across all three research questions.

TABLE II
DATASET CHARACTERISTICS

Property	CASIA-WebFace	DCFace
Type	Authentic	Synthetic
Generation	Web-crawled	Dual-cond. diffusion
Total images	0.5M	0.5M
Total identities	10K	10K
Images/identity	$\sim$116	50
Subset (ours)	10 ids / 1,165 imgs	10 ids / 550 imgs
Caucasian	63.4%	82.9%
Asian	14.4%	5.7%
African	7.4%	2.5%
Indian	7.2%	8.9%
Privacy	GDPR concerns	Privacy-friendly

TABLE III
FACE VERIFICATION BENCHMARKS

Benchmark	Full Name	Pairs	Variation
LFW [18]	Labeled Faces in the Wild	6,000	Unconstrained
CFP-FP [19]	Frontal-Profile	7,000	Cross-pose
CFP-FF [19]	Frontal-Frontal	7,000	Frontal pose
AgeDB-30 [20]	Age Database	6,000	Cross-age
CA-LFW [21]	Cross-Age LFW	6,000	Age-invariant
CP-LFW [22]	Cross-Pose LFW	6,000	Pose-invariant

B. Evaluation Benchmarks

Six standard face verification benchmarks are used, covering pose, age, and unconstrained real-world conditions (Table III). All benchmarks report verification accuracy (%) under their respective official protocols.

C. Loss Functions

Two margin-based loss functions are evaluated. CosFace [4] subtracts a fixed margin m from the target logit in the cosine domain:

$$\mathcal{L}_{\text{CosFace}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos\theta_{y_i}-m)}}{e^{s(\cos\theta_{y_i}-m)}+\sum_{j\neq y_i}^{c}e^{s\cos\theta_j}} \quad (1)$$

ArcFace [5] adds the margin m to the angle θ_{y_i} before applying the cosine:

$$\mathcal{L}_{\text{ArcFace}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)}+\sum_{j\neq y_i}^{c}e^{s\cos\theta_j}} \quad (2)$$

where θ_{y_i} is the angle between the feature embedding and its target class center, s is the scale factor, c the number of classes, and N the batch size. Both losses use s=64; m=0.35 for CosFace and m=0.50 for ArcFace.

D. Evaluation Metric

Verification accuracy (%) is computed following the standard protocol [18]:

$$\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives respectively. All results are reported as mean accuracy with standard deviation ($\pm$) over 10-fold cross-validation.

E. Training Configuration

All experiments are conducted on a single NVIDIA Tesla T4 GPU (15 GB VRAM). Models are optimized via SGD with a learning rate of 0.1, momentum of 0.9, and weight decay of 5×10^{-4}, over 2 epochs with a batch size of 8. Feature embeddings are projected to 128 dimensions. The loss scale is fixed at s=64 for both loss functions. All hyperparameters are kept strictly identical across data source, mixing ratio, backbone depth, and loss function configurations.

F. Backbone Architectures

Three residual network variants [13] are evaluated. ResNet34 uses Basic Blocks of two stacked 3×3 convolutions; ResNet50 and ResNet100 use Bottleneck Blocks with 1×1 reduction and expansion around a 3×3 convolution. The three backbones differ in depth and parameter count (Table IV).

V. RESULTS AND ANALYSIS

A. RQ1 – Authentic vs. Synthetic Data Across Backbone Depths

Table V reports verification accuracy for authentic (CASIA-WebFace) and synthetic (DCFace) single-source training across three backbone depths under CosFace loss.

TABLE IV
BACKBONE ARCHITECTURE COMPARISON

Property	ResNet34	ResNet50	ResNet100
Block type	Basic	Bottleneck	Bottleneck
Parameters	∼21M	∼25M	∼44M
Depth (layers)	34	50	100
FLOPs	∼3.6G	∼4.1G	∼7.8G
Complexity	Low	Medium	High

The accuracy gap between authentic and synthetic sources remains below 1% at every backbone depth, with no source consistently leading across all six benchmarks: DCFace averages 60.30% against 59.69% for CASIA-WebFace at ResNet34, while CASIA-WebFace leads by 0.47% and 0.75% at ResNet50 and ResNet100, all within the reported standard-deviation ranges. This comparison is conducted under a volume imbalance: the CASIA-WebFace subset contains 1,165 images per identity against 550 for DCFace — a 2.1× difference. The observed equivalence therefore reflects a joint effect of data source and volume and cannot be attributed to source alone; this confound is acknowledged as a limitation (Section VI-C). Average accuracy decreases consistently with backbone depth for both sources.

Answer to RQ1: Authentic and synthetic sources produce equivalent average accuracy at all three backbone depths (differences below 1%) despite a 2.1× image-volume advantage for CASIA-WebFace, suggesting DCFace is competitive under volume-constrained conditions, though this imbalance prevents a strictly controlled source-only comparison. Average accuracy decreases consistently with backbone depth for both sources.

B. RQ2 – Impact of Mixing Ratio Across Backbone Depths

Table VI reports verification accuracy as the number of synthetic identities increases from 0 to 30 in steps of 5, with 10 authentic identities fixed, across all three backbone depths under CosFace loss.

For ResNet34 and ResNet50, the highest average accuracy (60.24% and 60.37%) is reached at 5 synthetic identities (+0.55% and +0.62% over the 0-synthetic baseline), with no consistent gain beyond. For ResNet100, 5 synthetic identities bring almost no gain (+0.10%); the maximum of 59.33% is reached at 15 synthetic identities (+0.74% over baseline), beyond which accuracy fluctuates. Across all backbones, standard deviations decrease as the synthetic proportion grows, and CFP-FP shows a downward trend. Fig. 2 plots the average accuracy across mixing ratios.

Answer to RQ2: Mixed training improves over single-source training at all backbone depths. The optimal synthetic proportion is 5 identities for ResNet34 and ResNet50, and 15 identities for ResNet100. A paired t-test across the 10 evaluation folds confirms that the improvement of ResNet100 at 15 synthetic identities over the 0-synthetic baseline is statistically significant ($t = -4.83$, $p = 0.0009$).

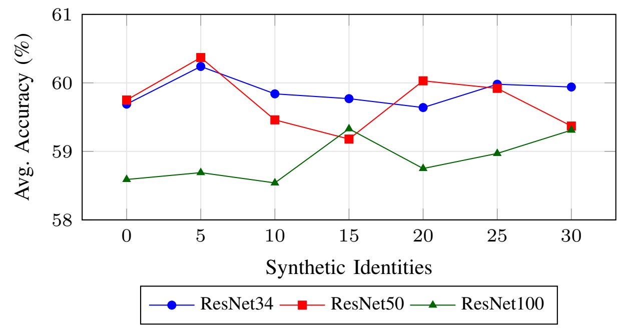

Fig. 2. Average verification accuracy vs. number of synthetic identities (10 authentic fixed) across backbone depths under CosFace loss.

C. RQ3 – Loss Function Interaction Across Data Sources and Backbone Depths

Building on the RQ2 optimal configurations, Table VII compares CosFace and ArcFace on authentic-only, synthetic-only, and combined data for each backbone depth (CosFace results carried over from RQ1 and RQ2).

On authentic-only data, CosFace and ArcFace are comparable at all depths (differences of 0.26%, 0.40%, and 0.07%). On synthetic-only data, CosFace outperforms ArcFace at all three depths, the gap shrinking from 1.25% (ResNet34) to 0.01% (ResNet100). On combined data, CosFace leads at ResNet34 (+0.31%) and ResNet50 (+0.88%), whereas at ResNet100 ArcFace records 60.54% against 59.66% for CosFace — the only configuration where ArcFace leads, driven by gains on LFW, CFP-FF, and AgeDB-30.

Answer to RQ3: CosFace records higher accuracy than ArcFace on synthetic-only data at all backbone depths, with the gap decreasing from 1.25% (ResNet34) to 0.01% (ResNet100). On combined data, CosFace leads at ResNet34 and ResNet50, while ArcFace leads at ResNet100 (60.54%), the highest average accuracy across all RQ3 configurations. However, a paired t-test across the 10 evaluation folds indicates that this difference does not reach statistical significance ($t = 0.02$, $p = 0.984$), suggesting the observed +0.88% gap may lie within the range of random variation under the low-data regime of this study. The relative performance of CosFace and ArcFace depends on both data configuration and backbone depth.

VI. DISCUSSION

A. Practical Recommendations

Based on the experimental results, the following guidelines are proposed:

- *Synthetic data proportion:* use 5 synthetic identities per 10 authentic for ResNet34 and ResNet50; use 15 for ResNet100.
- *Loss function selection:* use CosFace for single-source configurations at any backbone depth, and for shallow backbones on combined data; use ArcFace when training ResNet100 on combined authentic and synthetic data.
- *Efficiency trade-off:* under data scarcity, shallower backbones rival ResNet100. ResNet34 (CosFace, 5 syn-

TABLE V

VERIFICATION ACCURACY (%) — AUTHENTIC VS. SYNTHETIC DATA ACROSS BACKBONE DEPTHS (COSFACE)

Benchmark	ResNet34 CASIA	ResNet34 DCFace	ResNet50 CASIA	ResNet50 DCFace	ResNet100 CASIA	ResNet100 DCFace
LFW	69.13±3.01	70.60±2.16	70.75±2.71	69.37±2.12	67.90±2.27	67.30±1.83
CFP-FP	57.40±1.86	57.67±1.07	57.79±1.39	56.74±0.92	57.04±1.62	56.86±1.50
CFP-FF	70.64±2.49	69.46±1.97	67.70±1.91	68.11±1.65	68.27±2.13	66.03±2.41
AgeDB-30	49.87±1.85	52.62±2.89	50.67±1.40	50.83±0.65	50.73±2.20	49.38±1.61
CA-LFW	57.03±1.29	55.88±1.99	56.82±1.25	56.65±1.18	54.98±2.03	55.62±1.73
CP-LFW	54.08±1.75	55.45±1.68	54.77±1.83	52.95±1.46	52.67±1.63	51.87±1.53
Avg.	**59.69**	**60.30**	**59.75**	**59.28**	**58.59**	**57.84**

TABLE VI

VERIFICATION ACCURACY (%) VS. NUMBER OF SYNTHETIC IDENTITIES ACROSS BACKBONE DEPTHS (10 AUTHENTIC IDENTITIES FIXED, COSFACE)

Backbone	Synth	LFW	CFP-FP	CFP-FF	AgeDB	CA-LFW	CP-LFW	**Avg.**
ResNet34	0	69.13±3.01	57.40±1.86	70.64±2.49	49.87±1.85	57.03±1.29	54.08±1.75	59.69
	5	70.48±2.71	56.84±1.69	70.83±2.40	53.00±3.15	56.48±1.37	53.78±1.51	**60.24**
	10	69.68±1.54	57.10±1.10	70.06±2.35	50.80±2.28	57.42±1.59	54.00±1.52	59.84
	15	70.22±2.20	57.73±1.58	69.60±2.46	50.25±1.36	57.90±1.72	52.93±1.36	59.77
	20	69.85±1.77	56.89±1.52	68.47±2.35	52.98±2.24	56.90±1.69	52.77±0.91	59.64
	25	71.20±1.67	57.33±1.96	70.31±3.03	50.45±1.61	57.68±1.41	52.92±1.37	59.98
	30	70.38±2.42	56.73±1.54	69.69±1.37	51.52±1.85	58.25±1.29	54.08±1.40	59.94
ResNet50	0	70.75±2.71	57.79±1.39	67.70±1.91	50.67±1.40	56.82±1.25	54.77±1.83	59.75
	5	70.37±2.02	57.69±1.58	70.26±2.00	52.23±3.08	57.90±1.87	53.75±0.78	**60.37**
	10	69.87±1.85	57.39±1.40	70.66±2.20	49.73±1.93	56.33±1.11	52.77±1.79	59.46
	15	69.18±2.46	57.16±1.19	68.00±2.27	49.58±1.60	58.10±0.78	53.03±1.59	59.18
	20	70.82±1.95	57.57±1.52	70.01±2.00	51.38±1.73	57.75±1.35	52.65±0.80	60.03
	25	70.48±2.24	56.99±1.34	69.63±1.83	51.63±2.76	56.87±1.43	53.90±1.46	59.92
	30	71.48±2.10	55.69±1.12	69.93±1.65	49.60±1.70	57.32±1.26	52.20±1.15	59.37
ResNet100	0	67.90±2.27	57.04±1.62	68.27±2.13	50.73±2.20	54.98±2.03	52.67±1.63	58.59
	5	68.78±2.34	56.49±1.57	67.99±2.37	51.70±2.66	56.90±1.23	52.30±1.54	58.69
	10	68.37±2.68	55.94±1.66	68.90±1.94	49.03±1.23	56.37±1.58	52.65±1.35	58.54
	15	71.22±1.55	56.50±1.56	68.20±2.19	50.95±1.43	58.03±1.15	53.05±1.64	**59.33**
	20	70.83±1.98	55.93±0.69	66.07±1.99	49.17±1.90	57.07±1.71	53.43±1.16	58.75
	25	69.77±1.72	56.91±1.45	67.21±2.27	49.77±1.83	57.45±1.29	52.92±1.32	58.97
	30	69.57±1.69	56.27±1.12	67.96±2.43	51.62±1.78	57.28±1.09	53.13±1.73	59.31

TABLE VII

COSFACE VS. ARCFACE VERIFICATION ACCURACY (%) ACROSS DATA CONFIGURATIONS AND BACKBONE DEPTHS

Backbone	Benchmark	Auth only CosFace	Auth only ArcFace	Synth only CosFace	Synth only ArcFace	Combined* CosFace	Combined* ArcFace
ResNet34	LFW	69.13±3.01	69.13±2.94	70.60±2.16	68.95±2.27	70.48±2.71	69.67±2.21
	CFP-FP	57.40±1.86	57.26±1.80	57.67±1.07	55.97±1.22	56.84±1.69	56.94±1.65
	CFP-FF	70.64±2.49	68.77±2.34	69.46±1.97	68.56±2.04	70.83±2.40	69.90±2.14
	AgeDB-30	49.87±1.85	51.45±2.26	52.62±2.89	51.05±2.22	53.00±3.15	52.72±2.63
	CA-LFW	57.03±1.29	56.20±1.36	55.88±1.99	56.65±1.58	56.48±1.37	57.10±1.53
	CP-LFW	54.08±1.75	53.77±1.70	55.45±1.68	52.98±1.47	53.78±1.51	53.25±1.62
	Avg.	**59.69**	59.43	**60.28**	59.03	**60.24**	59.93
ResNet50	LFW	70.75±2.71	68.93±2.67	69.37±2.12	68.78±1.44	70.37±2.02	68.55±1.45
	CFP-FP	57.79±1.39	57.43±1.68	56.74±0.92	55.21±1.08	57.69±1.58	56.01±2.66
	CFP-FF	67.70±1.91	68.17±2.13	68.11±1.65	69.76±2.31	70.26±2.00	69.39±2.00
	AgeDB-30	50.67±1.40	52.38±2.43	50.83±0.65	49.75±1.67	52.23±3.08	53.50±1.44
	CA-LFW	56.82±1.25	55.83±1.42	56.65±1.18	56.55±1.25	57.90±1.87	57.70±1.40
	CP-LFW	54.77±1.83	53.38±1.71	52.95±1.46	51.10±1.14	53.75±0.78	51.77±0.93
	Avg.	**59.75**	59.35	**59.11**	58.52	**60.37**	59.49
ResNet100	LFW	67.90±2.27	69.08±2.33	67.30±1.83	68.23±1.96	71.22±1.55	71.83±2.77
	CFP-FP	57.04±1.62	56.84±1.38	56.86±1.50	54.36±1.39	56.50±1.56	57.20±1.28
	CFP-FF	68.27±2.13	66.74±2.56	66.03±2.41	65.41±1.75	68.20±2.19	71.04±1.99
	AgeDB-30	50.73±2.20	50.10±0.51	49.38±1.61	49.40±1.01	50.95±1.43	52.22±2.30
	CA-LFW	54.98±2.03	55.08±1.34	55.62±1.73	57.25±1.41	58.03±1.15	58.05±1.19
	CP-LFW	52.67±1.63	54.15±1.78	51.87±1.53	52.32±1.03	53.05±1.64	52.92±1.91
	Avg.	58.60	**58.66**	**57.84**	57.83	59.66	**60.54**

* 10+5 for ResNet34/ResNet50, 10+15 for ResNet100.

thetic) reaches 60.24% against 60.54% for ResNet100 (ArcFace, 15 synthetic) at far lower cost (3.6G vs. 7.8G FLOPs, Table IV); the 0.30% gain may not justify the 2.2× inference cost in resource-constrained applications.

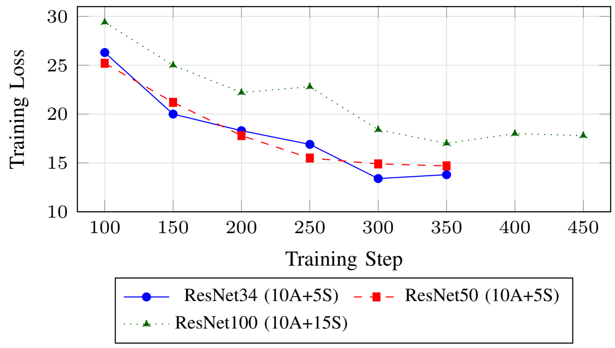

Fig. 3. Training loss over two epochs for the three optimal configurations under CosFace loss.

B. Relationship to Prior Work

The competitive accuracy of DCFace relative to CASIA-WebFace agrees with Kim *et al.* [6], and the accuracy plateau beyond the optimal mixing ratio matches the saturation effect of DigiFace-1M [12]. The loss-data interaction in RQ3 extends Atzori *et al.* [11], who used CosFace on a single backbone, by showing that the CosFace-ArcFace ranking varies with both data configuration and backbone depth. The accuracy levels here (best: 60.54%) are well below state-of-the-art ArcFace on full CASIA-WebFace (99.83% on LFW [5]); this gap is intentional, as the constrained 10–40 identity regime isolates the joint effect of backbone depth, mixing ratio, and loss rather than maximizing absolute accuracy.

C. Limitations and Future Work

This study has five limitations. First, experiments use small subsets of 10 to 40 identities per source, and results may not generalize to full-scale datasets. Second, the RQ1 comparison is conducted under an uncontrolled image-volume imbalance (1,165 vs. 550 images per identity, a $2.1\times$ difference); a fully controlled comparison would require subsampling CASIA-WebFace to 550 images per identity, left for future work. Third, only CosFace and ArcFace are evaluated; AdaFace [17] is not included. Fourth, the training schedule is fixed at two epochs with a constant learning rate, which prevents the learning-rate scheduler from activating and may limit deeper backbones; nonetheless, all three backbones show a consistent decrease in training loss (Fig. 3), and ResNet100, despite a higher initial loss from its larger parameter space, reaches a comparable final trajectory, indicating that the fixed schedule does not prevent learning even for the deepest backbone.

Fifth, evaluation is restricted to verification accuracy; fairness across demographic groups is not measured. Future work should validate these results at larger training scales, evaluate AdaFace as an additional loss function, extend the backbone comparison beyond ResNet, and assess fairness across demographic groups using ethnicity-stratified benchmarks.

VII. CONCLUSION

This paper studied the joint effect of synthetic data proportion, backbone depth, and loss function on face verification accuracy, using unbalanced subsets of CASIA-WebFace and DCFace across three ResNet depths and two margin-based losses on six benchmarks. Three results emerge. First, synthetic and authentic data yield comparable accuracy across all depths (within 1%), confirming DCFace as a privacy-friendly substitute. Second, mixing improves over single-source training, but the optimal synthetic proportion varies with depth: 5 identities for ResNet34 and ResNet50, 15 for ResNet100. Third, CosFace outperforms ArcFace on single-source data at all depths, while ArcFace leads on combined data for ResNet100 (60.54% vs. 59.66%), showing the best loss depends on both data configuration and depth. Backbone depth, mixing ratio, and loss therefore interact and should be configured jointly. Future work should validate these findings at larger scale, include AdaFace [17], and assess fairness across demographic groups.

REFERENCES

[1] European Parliament, "Regulation (EU) 2016/679 (General Data Protection Regulation)," *Official Journal of the European Union*, 2016.
[2] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv:1411.7923*, 2014.
[3] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *IEEE FG*, 2018, pp. 67–74.
[4] H. Wang *et al.*, "CosFace: Large margin cosine loss for deep face recognition," in *CVPR*, 2018, pp. 5265–5274.
[5] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE TPAMI*, vol. 44, no. 10, pp. 5962–5979, 2022.
[6] M. Kim, F. Liu, A. K. Jain, and X. Liu, "DCFace: Synthetic face generation with dual condition diffusion model," in *CVPR*, 2023, pp. 12715–12725.
[7] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer, "IDiff-Face: Synthetic-based face recognition through fuzzy identity-conditioned diffusion model," in *ICCV*, 2023, pp. 19650–19661.
[8] F. Boutros, V. Štruc, J. Fierrez, and N. Damer, "Synthetic data for face recognition: Current state and future prospects," *Image and Vision Computing*, vol. 135, p. 104688, 2023.
[9] M. Wang and W. Deng, "Mitigating bias in face recognition using skewness-aware reinforcement learning," in *CVPR*, 2020, pp. 9322–9331.
[10] J. J. Howard, Y. B. Sirotin, and A. R. Vemury, "The effect of broad and specific demographic homogeneity on the imposter distributions and false match rates in face recognition algorithm performance," in *IEEE BTAS*, 2019, pp. 1–8.
[11] A. Atzori, P. Cosseddu, G. Fenu, and M. Marras, "The impact of balancing real and synthetic data on accuracy and fairness in face recognition," in *ECCV Workshops*, 2024.
[12] G. Bae *et al.*, "DigiFace-1M: 1 million digital face images for face recognition," in *IEEE WACV*, 2023, pp. 3515–3524.
[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.
[14] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao, "SynFace: Face recognition with synthetic data," in *ICCV*, 2021, pp. 10860–10870.
[15] F. Boutros, M. Huber, P. Siebke, T. Rieber, and N. Damer, "SFace: Privacy-friendly and accurate face recognition using synthetic data," in *IEEE IJCB*, 2022, pp. 1–11.
[16] F. Boutros, M. Klemt, M. Fang, A. Kuijper, and N. Damer, "ExFaceGAN: Exploring identity directions in GAN's learned latent space for synthetic identity generation," in *IEEE IJCB*, 2023.
[17] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *CVPR*, 2022, pp. 18750–18759.
[18] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild," in *Workshop on Faces in Real-Life Images*, 2008.
[19] S. Sengupta *et al.*, "Frontal to profile face verification in the wild," in *IEEE WACV*, 2016, pp. 1–9.
[20] S. Moschoglou *et al.*, "AgeDB: The first manually collected in-the-wild age database," in *CVPRW*, 2017, pp. 51–59.
[21] T. Zheng, W. Deng, and J. Hu, "Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments," *arXiv:1708.08197*, 2017.
[22] T. Zheng and W. Deng, "Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments," BUPT, Tech. Rep., 2018.