Domain Balancing: Face Recognition on Long-Tailed Domains

Dong Cao[1,2*] Xiangyu Zhu[1,2*] Xingyu Huang[3] Jianzhu Guo[1,2] Zhen Lei[1,2†]
[1]CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
[3]Tianjin University
{dong.cao,xiangyu.zhu,jianzhu.guo,zlei}@nlpr.ia.ac.cn, xingyu.huang@tju.edu.cn

Abstract

Long-tailed problem has been an important topic in face recognition task. However, existing methods only concentrate on the long-tailed distribution of classes. Differently, we devote to the long-tailed domain distribution problem, which refers to the fact that a small number of domains frequently appear while other domains far less existing. The key challenge of the problem is that domain labels are too complicated (related to race, age, pose, illumination, etc.) and inaccessible in real applications. In this paper, we propose a novel Domain Balancing (DB) mechanism to handle this problem. Specifically, we first propose a Domain Frequency Indicator (DFI) to judge whether a sample is from head domains or tail domains. Secondly, we formulate a light-weighted Residual Balancing Mapping (RBM) block to balance the domain distribution by adjusting the network according to DFI. Finally, we propose a Domain Balancing Margin (DBM) in the loss function to further optimize the feature space of the tail domains to improve generalization. Extensive analysis and experiments on several face recognition benchmarks demonstrate that the proposed method effectively enhances the generalization capacities and achieves superior performance.

1. Introduction

Feature descriptor is of crucial importance to the performance of face recognition, where the training and testing images are drawn from different identities and the distance metric is directly acted on the features to determine whether they belong to the same identity or not. Recent years have witnessed remarkable progresses in face recognition, with a variety of approaches proposed in the literatures and applied in real applications [18, 32, 4, 7, 6, 42]. Although yielding excellent success, face recognition often suffers from poor generalization, i.e., the learned features only work well on the domain the same as the training set and perform poorly on the unseen domains. This is one of the most critical issues for face recognition in the wild, partially due to the non-negligible domain shift from the training set to the deployment environment.

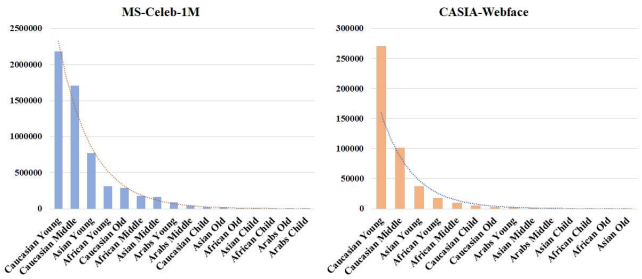

Figure 1. The long-tailed domain distribution demarcated by the mixed attributions (e.g., race and age) in the MS-Celeb-1M [8] and CASIA-Webface [36]. Number of classes per domain falls drastically, and only few domains have abundant classes. (Baidu API [1] is used to estimate the race and age)

Real-world visual data inherently follows a long-tailed distribution, where only a limited number of classes appear frequently, and most of the others remain relatively rare. In this paper, we aim to investigate the long-tailed domain distribution and balance it to improve the generalization capacity of deep models. However, different from the long-tailed problem in classes, domain labels are inaccessible in most of applications. Specifically, domain is an abstract attribute related to many aspects, e.g., age (baby, child, young man, aged, etc), race (caucasian, indian, asian, african, etc.), expression (happy, angry, surprise, etc.), pose (front, profile, etc.), etc. As a result, the domain information is hard to label or even describe. Without the domain label, it is difficult to judge whether a sample belongs to the head domains or the tail domains, making existing methods inapplicable. Figure 1 illustrates a possible partition by the mixed attributions (e.g., race and age).

We formally study this long-tailed domain distribution

*Equally-contributed
†Corresponding author

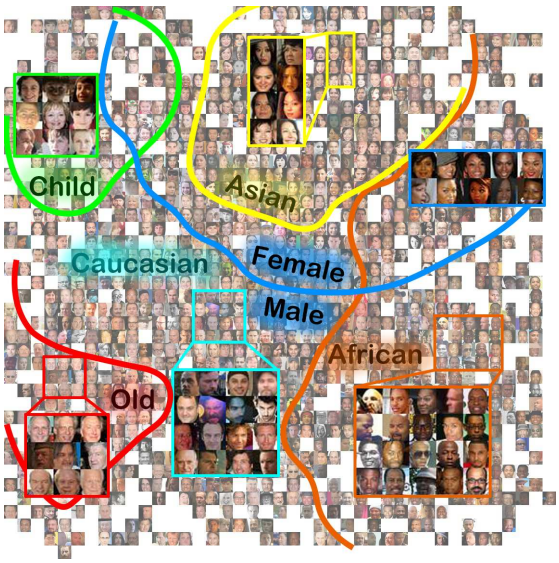

Figure 2. The face features are trivially grouped together according to different attributions, visualized by t-SNE [20]

problem arising in real-world data settings. Empirically, the feature learning process will be significantly dominated by those few head domains while ignoring many other tail domains, which increases the recognition difficulty in the tail domains. Such undesirable bias property poses a significant challenge for face recognition systems, which are not restricted to any specific domain. Therefore, it is necessary to enhance the face recognition performance regardless of domains. An intuitive method to handle the long-tailed problem is over-sampling and under-sampling samples on the tail and the head, respectively [10, 27, 41]. However, it does not work well on domain balancing since the the ground-truth domain distribution is inaccessible. To overcome this drawback, we propose a Domain Balancing (DB) mechanism to balance the long-tailed domain distribution.

Firstly, since the ground truth domain distributions are inaccessible without domain labels, for each sample, we should predict where the belonged domain locates on the distribution. To this end, we point out that the domain frequency can be instructed by the inter-class compactness. In the holistic feature space, the classes with similar attributes tend to group, forming a specific domain as shown in Figure 2. Besides, in the feature space, the compactness is not everywhere equal. Take the domains in Figure 1 as an example, we find the compact regions tend to belong to the head domains (e.g., caucasian male), and the sparse regions tend to belong to the tail domains (e.g., children, african female). The detailed analysis will be shown in Section 3.1. Motivated by these observations, we propose to utilize the inter-class compactness which is the local distances within a local region as the Domain Frequency Indicator (DFI). Secondly, considering the samples belong to the same domain share some appearance consistency, we design a novel module called Residual Balancing Mapping (RBM) block, which can adaptively change the network based on DFI to find the best network to adjust each domain. The block consists of two components: a domain enhancement branch and a soft gate. The domain enhancement branch aims to adjust the network to each domain through enhancement residual and the soft gate attaches a harmonizing parameter to the residual to control the amount of residual according to the domain frequency. Thirdly, in the loss function, we propose a Domain Balancing Margin (DBM) to adaptively modify the margin according to the DFI for each class, so that the loss produced by the tail domain classes can be relatively up-weighted. The framework is shown in Figure 3.

The major contributions can be summarized as follows:

- We highlight the challenging long-tailed domain problem, where we must balance the domain distribution without any domain annotation.
- We propose a Domain Balancing (DB) mechanism to solve the long-tailed domain distribution problem. The DB can automatically evaluate the domain frequency of each class with a Domain Frequency Indicator (DFI) and adapt the network and loss function with Residual Balancing Mapping (RBM) and Domain Balancing Margin (DBM), respectively.
- We evaluate our method on several large-scale face datasets. Experimental results show that the proposed Domain Balancing can efficiently mitigate the long-tailed domain distribution problem and outperforms the state-of-the-art approaches.

2. Related Works

Softmax based Face Recognition. Deep convolutional neural networks (CNNs) [3] have achieved impressive success in face recognition. The current prevailing softmax loss considers the training process as a N-way classification problem. Sun et al. [28] propose the DeepID for face verification. In the training process, for each sample, the extracted feature is taken to calculate the dot products with all the class-specific weights. Wen et al. [35] propose a new center loss penalizing the distances between the features and their corresponding class centers. Wang et al. [31] study the effect of normalization during training and show that optimizing cosine similarity (cosine-based softmax loss) instead of inner-product improves the performance. Recently, a variety of margin based softmax losses [18, 32, 4] have achieved the state-of-the-art performances. SphereFace [18] adds an extra angular margin to attain shaper decision boundary of the original softmax loss. It concentrates the features in a sphere mainfold. CosFace [32] shares a similar idea which encourages the

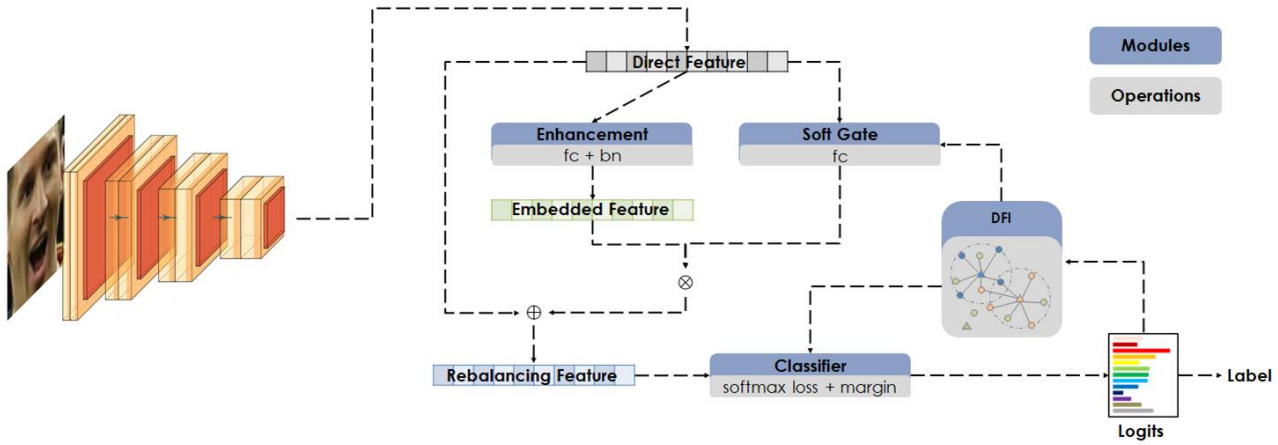

Figure 3. There are three main modules: DFI, RBM and DBM. The DFI indicates the local distances within a local region. The RBM harmonizes the representation ability in the network architecture, while the DBM balances the contribution in the loss.

intra-compactness in the cosine manifold. Another effort ArcFace [4] uses an additive angular margin, leading to similar effect. However, these efforts only consider the intra-compactness. RegularFace [38] proposes an exclusive regularization to focus on the inter-separability. These methods mainly devote to enlarge the inter-differences and reduce the intra-variations. Despite their excellent performance on face recognition, they rely more on the large and balanced datasets and often suffer performance degradation when facing with the long-tailed data.

Long – tailed Learning Long-tailed distribution of data has been well studied in [37, 19]. Most existing methods define the long-tailed distribution in term of the size of each class. A widespread method is to resample and rebalance training data, either by under-sampling examples from the head data [10], or over-sampling samples from the rare data more frequently [27, 41]. The former generally loses critical information in the head sets, whereas the latter generates redundancy and may easily encounter the problem of over-fitting to the rare classes. Some recent strategies include hard negative mining [5, 15], metric learning [12, 23] and meta learning [9, 34]. The range loss [37] proposes an extra range constraint jointly with the softmax loss. It reduces the k greatest intra-class ranges and enlarges the shortest inter-class distance within one batch. The focal loss [15] employs an online version of hard negative mining. Liu et al. [19] investigate the long-tailed problem in the open set. Its so-called dynamic meta-embedding uses an associated memory to enhance the representation. Adaptiveface [16] analyzes the difference between rich and poor classes and proposes the adaptive margin softmax to dynamically modify the margins for different classes. Although the long-tailed problem has been well studied, they are mainly based on the category frequency distribution. None of previous works consider the similar problem in domain. One possible reason may be due to the ambiguous domain partition as discussed above. In fact, the domains may not even have explicit semantics, i.e., they are actually data-driven.

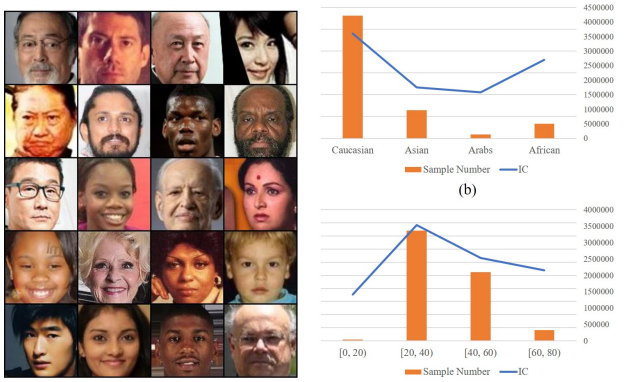

Figure 4. (a) Identities with small inter-class compactness value in the MS-Celeb-1M. (b) The inter-class compactness vs. race distribution. (c) The inter-class compactness vs. age distribution.

In contrast, our method focuses on the long-tailed domain, which is more in line with the real-world application. The proposed method balances the contribution of domains on the basis of their frequency distribution, so that it can improve the poor generalization well.

3. Domain Balancing

We propose to balance the samples from different domains without any domain annotation. Domain Balancing (DB) mechanism has three components: Domain Frequency Indicator (DFI) to evaluate the domain frequency, the Residual Balancing Mapping (RBM) to adjust the network and the Domain Balancing Margin (DBM) to adjust the loss functions according to domain distribution.

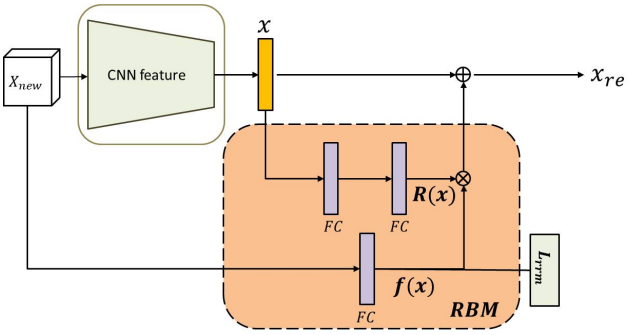

Figure 5. The Residual Balancing Module is designed with light-weighted structure and it can be easily attached to existing network architecture. The block dynamically enhances the feature according to DFI.

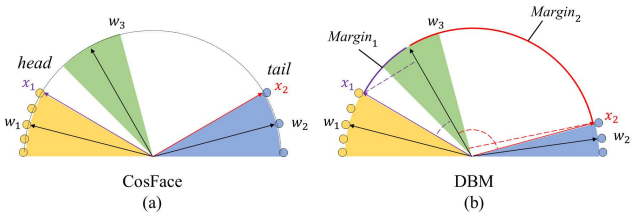

Figure 6. Geometrical interpretation of DBM from the feature perspective. Different color areas indicate feature space from distinct classes. Yellow area represents the head-domain class C_1 and blue area represents the tailed-domain class C_2. (a) CosFace assigns an uniform margin for all the classes. The sparse inter-class distribution in the tail domains makes the decision boundary easy to satisfy. (b) DBM assigns margin according to the inter-class compactness adaptively.

3.1. Domain Frequency Indicator

To handle the long-tailed domain distribution problem, we first need to know whether a sample is from a head domain or from a tail domain. We introduce a Domain Frequency Indicator (DFI) based on the inter-class compactness. Inter-class compactness function of a given class is formulated as:

$$IC(w) = log \sum_{k=1}^{K} e^{s \cdot cos(w_k, w)} \tag{1}$$

where w is the prototype of one class in the classification layer and k is the k-th nearest class, where the distance of two classes i, j is formulated as $cos(w_i, w_j)$. The high frequency domain, i.e., head domain, usually corresponds to a large $IC(w)$, and vice versa. Then we define the Domain Frequency Indicator as:

$$DFI = \frac{\varepsilon}{IC(w)} \tag{2}$$

which is inversely proportional to the inter-class compactness $IC(w)$ and ε is a constant value. Ideally, if the classes are uniformly distributed, each class will have the same DFI. Otherwise, the classes with larger DFI are more likely to come from a tail domain and should be relatively up-weighted. As shown in Figure 4, the identities with larger DFI values usually come from Africa, children or the aged, which are highly related with the tail domains.

3.2. Residual Balancing Module

In real-world application, face recognition accuracy depends heavily on the quality of the top-level feature x. The goal in this section is to design a light-weight solution to adjust the network to extract domain specific features according to the domain distribution. Our Residual Balancing Module (RBM) combines the top-level image feature and a residual feature, using DFI to harmonize the magnitude.

Even though big training data facilitates the feature discriminative power, the head domains dominate the learning process and the model lacks adequate supervised updates from the tail classes. We hope to learn a harmonizing module through a mapping function $M_{re}(.)$ to adjust the features for samples of different domain frequency to mitigate the domain imbalance problem. We formulate M_{re} as a sum of the original feature x and residual acquired by a feature enhancement module $R(x)$ weighted by $f(x)$. We denote the resulting feature as x_{re} and the RRM can be formulated as:

$$\begin{aligned} x_{re} &= M_{re}(x) \\ &= x + f(x) \cdot R(x) \end{aligned} \tag{3}$$

where x is the top-level feature, $f(x)$ is a soft gate depending on the DFI. When DFI is large, the input feature probably belongs to a tail class, and a large enhancement is assigned to the residual. Otherwise, the enhancement is trivial. The magnitude of residual is thus inversely proportional to the domain frequency. The combination of the soft gate and and the residual can be regarded as a harmonizing mechanism that adopts domain distribution information to control the magnitude to be passed to the next layer.

We now describe the implementation of the two components: The first component is the residual $R(x)$, which is implemented by a light-weighted full-connected layer. It consists of two full-connected layers and a batch norm layer shown in Figure 5. The second component is the soft gate coefficient DFI, which is learned from the feature x and supervised by the DFI. For simplicity, the linear regression is employed by the L2 loss:

$$L_{rrm} = \|f(x) - DFI(x)\|_2^2 \tag{4}$$

where $DFI(x)$ is defined in Eq. 2 from the last iteration. $f(x)$ is a mapping function devoting to associate the representation x and DFI.

3.3. Domain Balancing Margin

We propose a domain-aware loss by Domain Balancing Margin (DBM) to adaptively strengthen the classes in the tail domains. Specifically, we formulate the DBM loss by embedding the DFI into CosFace as:

$$L_{dbm} = -logP_{i,y_i} = -log\frac{e^{s(cos\theta_{i,y_i}-\beta_{y_i}\cdot m)}}{e^{s(cos\theta_{i,y_i}-\beta_{y_i}\cdot m)}+\sum_{k\neq y_i}^{C} e^{s\cdot cos\theta_{i,k}}} \tag{5}$$

where $\beta_{y_i} = DFI_{y_i}$ and m is a fixed parameter as defined in CosFace. Figure 6 visualizes the phenomenon through a trinary classification. The main difference between DBM and CosFace is that our margin is dynamic and feature compactness related. For the CosFace, the decision boundary assigns the same margin without considering the feature compactness. It cannot efficiently compact the feature space of the tailed-domain class C_2 since the sparse inter-class distribution makes the decision boundary easy to satisfy. The termination of optimization is so early, leading to poor generalization. In contrast, our DBM drives adaptive decision boundary in terms of the inter-compactness, where $margin_2$ (tailed-domain margin) should be much larger than $margin_1$ (head-domain margin). Consequently, both the inter-separability and the intra-compactness can be guaranteed.

We combine the mentioned L_{dbm} and L_{rrm} by a parameter λ. The final loss function can be formulated as:

$$L = L_{dbm} + \lambda L_{rrm} \tag{6}$$

4. Experiments

4.1. Datasets

Training Set. We employ CASIA-Webface [36] and MS-Celeb-1M [8] as our training sets. CASIA-WebFace is collected from the web. The face images are collected from various professions and suffer from large variations in illumination, age and pose. MS-Celeb-1M is one of the largest real-world face datasets containing 98,685 celebrities and 10 million images. Considering the amount of noise, we use a refined version called MS1MV2 [4] where a lot of manual annotations are employed to guarantee the quality of the dataset.

Testing Set. During testing, we firstly explore databases (RFW [33], AFW [2]) with obvious domain bias to check the improvement. RFW is a popular benchmark for racial bias testing, which contains four subsets, Caucasian, Asian, India and African. Moreover, we collect a new dataset from CACD [2], called Age Face in-the-Wild (AFW). We construct three testing subsets, Young (14-30 years old), Middle-aged (31-60 years old) and Aged (60-90 years old). Each subset contains 3,000 positive pairs and 3,000 negative pairs respectively. Besides, we further report the performance on several widely used benchmarks including LFW [13], CALFW [40], CPLFW [39] and AgeDB [21]. LFW contains color face images from 5,749 different persons in the web. We verify the performance on 6,000 image pairs following the standard protocol of unrestricted with labeled outside data. CALFW is collected with obvious age gap to add aging process intra-variance on the Internet. Similarly, CPLFW is collected in terms of pose difference. AgeDB contains face images from 3 to 101 years old. We use the most challenging subset AgeDB-30 in the following experiments. We also extensively evaluate our proposed method on large-scale face dataset, MegaFace [14]. MegaFace is one of the most challenging benchmark for large scale face identification and verification. The gallery set in MegaFace includes 1M samples from 690K individuals and the probe set contains more than 100K images of 530 different individuals from FaceScrub [22]. Table 1 shows the detailed information of the involved datasets.

Table 1. Statistics of face datasets for training and testing. (P) and (G) indicates the probe and gallery set respectively.

	Dataset	Identities	Images
Training	CASIA [36]	10K	0.5M
	MS1MV2 [4]	85K	5.8M
Testing	LFW [13]	5749	13,233
	CPLFW [39]	5,749	12,174
	CALFW [40]	5,749	11,652
	AgeDB [21]	568	16,488
	RFW [33]	11,430	40,607
	CACD [2]	2,000	160,000
	MegaFace [14]	530 (P)	1M(G)

4.2. Experimental Settings

For data prepossessing, the face images are resized to 112×112 by employing five facial points, and each pixel in RGB images is normalized by subtracting 127.5 and dividing by 128. For all the training data, only horizontal flipping is used for data augmentation. For the embedding neural network, we employ the widely used CNNs architectures, ResNet18 and ResNet50 [11]. They both contain four residual blocks and finally produce a 512-dimension feature.

In all the experiments, the CNNs models are trained with stochastic gradient descents (SGD). We set the weight decay of 0.0005 and the momentum of 0.9. The initial learning rate starts from 0.1 and is divided by 10 at the 5, 8, 11 epochs. The training process is finished at 15-th epoch. We set $\varepsilon = 5.5$ and $\lambda = 0.01$ in all the experiments. The experiments are implemented by PyTorch [25] on NVIDIA Tesla V100 (32G). We train the CNNs models from scratch and only keep the feature embedding part without the final fully connected layer (512-D) during testing.

Table 2. Face verification results (%) with different strategies. (CASIA-Webface, ResNet18, RBM (w/o sg) refers to RBM without the soft gate, i.e., $f(x) = 1$.)

	Module			LFW	CALFW	CPLFW	AgeDB	Average
	RBM(w/o sg)	RBM	DBM					
CASIA-Webface				98.8	91.0	85.4	90.2	91.35
			✓	99.2	92.0	87.3	91.9	92.6
		✓		99.1	91.0	87.1	91.3	92.12
	✓			98.7	90.6	85.4	90.3	91.25
		✓	✓	99.3	92.5	87.6	92.1	92.88

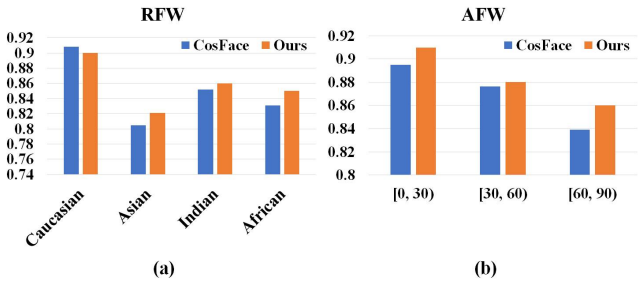

Figure 7. (a) The performance on four testing subsets, Caucasian, Indian, Asian and African in RFW. (b) The performance on three testing subsets, Young [0-30), Middle-aged [30-60) and Aged [60-90) in AFW.

We use cosine distance to calculate the similarity. For the performance evaluation, we follow the standard protocol of unrestricted with labeled outside data [13] to report the performance on LFW, CALFW, CPLFW, AgeDB, RFW and AFW. Considering the well solved on LFW, we further use the more challenging LFW BLUFR protocol to evaluate the proposed method. On MegaFace, there are two challenges. We use large protocol in Challenge 1 to evaluate the performance of our approach. For the fair comparison, we also clean the noisy images in Face Scrub and MegaFace by the noisy list [4].

To the compared approaches, we compare the proposed method with the baseline Softmax loss and the recently popular state-of-the-arts, including SphereFace [18], CosFace [32] and ArcFace [4].

4.3. Ablation Study

In this section, we investigate the effectiveness of each balancing module in the proposed method.

Effectiveness of the RBM. Recall that the RBM module consists of two main components: the residual enhancement and the soft gate. The soft gate produces a harmonizing coefficient to automatically control the magnitude of the residual attached to the top feature. When the soft gate is closed, i.e, $f(x) = 1$ is constant for all samples, the RBM module degenerates to a conventional residual that loses the ability of distinguishing the head and tail domains. From Table 2, we observe that the combination of the residual enhancement and the soft gate brings large improvements on all the datasets. The average performance of LFW, CALFW, CPLFW, AgeDB has been improved from 91.35 to 92.12. It is because RBM actually harmonizes the potential feature bias among different domains.

Effectiveness of the Soft Gate. The soft gate produces the coefficient DFI to control the magnitude of residual added on the original feature. In this experiment we analyze the effectiveness of the soft gate. As displayed in Table 2, the performance drops significantly without the soft gate. The average accuracy decreases 0.87%. These results suggest that the improvement attained by the RBM block is not mainly due to the additional parameters, but its internal domain balancing mechanism.

Effectiveness of the DBM. We further validate the effectiveness of DBM that whether it can improve the poor generalization caused by the long-tailed domain distribution. From the first row of each sub-boxes in Table 2, we can find that DBM boosts the performance on all the datasets. The average performance is stably improved compared to the baselines, presenting its contribution to mitigate the potential imbalance. Particularly, DBM achieves about 0.48% average improvement over RBM, which indicates that balancing the contribution from different domains through loss function can better address the problem.

4.4. Exploratory Experiments

We first investigate how our method improves the performance on the different domains with different domain frequency. We train Resnet50 on CASIA-Webface by CosFace and our method. Figure 7 shows the performances on different domains on two datasets. Firstly, for the Cosface, the accuracy of Caucasian on RFW is significantly higher than other races, and Asian gains the worse performance. Besides, on AFW, the Young subset acquires the highest accuracy while the performance on the aged persons degrades heavily. The performance decay confirms our thought that the popular methods is susceptible to the long-tailed domain distribution. Secondly, our method consistently improves the performance on almost all the domains. Particularly, the accuracy increases more obviously on the tail domains, such as the Asian on RFW and $[60, 90)$ aged persons on AFW,

which indicates that the proposed method can alleviate the potential imbalance cross domains.

The nearest neighbor parameter K in Eq. 1 plays an important role in DFI. In this part we conduct an experiment to analyze the effect of K. We use CASIA-WebFace and ResNet18 to train the model with our method and evaluate the performance on the LFW, CALFW and CPLFW as presented in Table 3. We can conclude that the model without DFI suffers from the poor performances on all these three benchmarks. The model attains the worst result on all the datasets when $K = 0$, where the model degenerates into the original form without balancing representation and margin supplied by RBM and DBM. The model obtains the highest accuracy at $K = 100$. However, when K keeps increasing, the performances decrease to some extent because a too large K covers a too large region with sufficient samples and weakens the difference between head and tail domain.

Table 3. Performance (%) vs. K on LFW, CALFW and CPLFW datasets, where K is the number of nearest neighbor in Domain Frequency Indicator (DFI).

K	0	100	1,000	3,000	6,000
LFW	98.8	99.3	99.1	99.2	99.2
CALFW	91.0	92.5	92.1	92.2	92.1
CPLFW	85.4	87.6	87.2	87.3	87.3

4.5. Evaluation Results

4.5.1 Results on LFW and LFW BLUFR

LFW is the most widely used benchmark for unconstrained face recognition. We use the common larget dataset MSIMV2 to train a ResNet50. Table 4 displays the the comparsion of all the methods on LFW testset. The proposed method improves the performance from 99.62% to 99.78%. Further, we evaluate our method on the more challenge LFW BLUFR protocol. The results are reported in Table 5. Despite the limited improvement, our approach still achieves the best results compared to the state-of-the-arts.

4.5.2 Results on CALFW, CPLFW and AgeDB

Table 6 shows the performances on CALFW, CPLFW and AgeDB, respectively. We also use MSIMV2 to train the ResNet50. The results show the similar treads that emerged on the previous test sets. Particularly, the margin-based methods attain better results than the simple softmax loss for face recognition. Our proposed method, containing efficient domain balancing mechanism, outperforms all the other methods on these three datasets. Specifically, our method achieves 95.54% average accuracy, about 0.4% average improvement over ArcFace.

Table 4. Face verification (%) on the LFW dataset. "Training Data" indicates the size of the training data involved. "Models" indicates the number of models used for evaluation.

Method	Training Data	Models	LFW
Deep Face [30]	4M	3	97.35
FaceNet [26]	200M	1	99.63
DeepFR [24]	2.6M	1	98.95
DeepID2+ [29]	300K	25	99.47
Center Face [35]	0.7M	1	99.28
Baidu [17]	1.3M	1	99.13
Softmax	5M	1	99.43
SphereFace [18]	5M	1	99.57
CosFace [32]	5M	1	99.62
ArcFace [4]	5M	1	99.68
Ours	5M	1	**99.78**

Table 5. Face verification (%) on LFW BLUFR protocol.

Method	VR@FAR =0.001%	VR@FAR =0.01%
Softmax	87.53	93.03
SphereFace [18]	98.50	99.17
CosFace [32]	98.70	99.20
ArcFace [4]	98.77	99.23
Ours	**98.91**	**99.53**

Table 6. Face verification (%) on CALFW, CPLFW and AgeDB.

Method	CALFW	CPLFW	AgeDB
Softmax	89.41	81.13	94.77
SphereFace [18]	90.30	81.40	97.30
CosFace [32]	93.28	92.06	97.70
ArcFace [4]	95.45	92.08	97.83
Ours	**96.08**	**92.63**	**97.90**

4.5.3 Results on MegaFace

We also evaluate our method on the large Megaface testset. Table 7 displays the identification and verification performances. In particular, the proposed method surpasses the best approach ArcFace by an obvious margin (about 0.82% at Rank-1 identification rate and 0.68% verification rate). The reason behind may be that the proposed balancing strategy can efficiently mitigate the potential impact of the long-tailed domain distribution, which is ubiquitous in the real-world application.

5. Conclusion

In this paper, we investigate a novel long-tailed domain problem in the real-world face recognition, which refers to few common domains and many more rare do-

Table 7. Face identification and verification on MegaFace Challenge1. "Rank 1" refers to the rank-1 face identification accuracy, and "Ver" refers to the face verification TAR at 10^{-6} FAR.

Method	Rank1 (%)	Ver (%)
DeepSense V2	81.29	95.99
YouTu Lab	83.29	91.34
Vocord-deepVo V3	91.76	94.96
SphereFace [18]	92.05	92.42
CosFace [32]	94.84	95.12
ArcFace [4]	95.53	95.88
Ours	**96.35**	**96.56**

mains. A novel Domain Balancing mechanism is proposed to deal with this problem, which contains three components, Domain Frequency Indicator (DFI), Residual Balancing Mapping (RBM) and Domain Balancing Margin (DBM). Specifically, DFI is employed to judge whether a class belongs to a head domain or a tail domain. RBM introduces a light-weighted residual controlled by the soft gate. DBM assigns an adaptive margin to balance the contribution from different domains. Extensive analyses and experiments on several face recognition benchmarks demonstrate that the proposed method can effectively enhance the discrimination and achieve superior accuracy.

Acknowledgement

This work has been partially supported by the Chinese National Natural Science Foundation Projects #61876178, #61806196, #61976229, #61872367

References

[1] Baidu cloud vision api. http://ai.baidu.com.

[2] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European conference on computer vision*, pages 768–783. Springer, 2014.

[3] Y Le Cun. Convolutional networks for images, speech, and time series. 1995.

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[5] Qi Dong, Shaogang Gong, and Xiatian Zhu. Class rectification hard mining for imbalanced deep learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1851–1860, 2017.

[6] Jianzhu Guo, Xiangyu Zhu, Zhen Lei, and Stan Z Li. Face synthesis for eyeglass-robust face recognition. In *Chinese Conference on Biometric Recognition*, pages 275–284. Springer, 2018.

[7] Jianzhu Guo, Xiangyu Zhu, Chenxu Zhao, Dong Cao, Zhen Lei, and Stan Z Li. Learning meta face recognition in unseen domains. *arXiv preprint arXiv:2003.07733*, 2020.

[8] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016.

[9] David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

[10] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284, 2009.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[12] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5375–5384, 2016.

[13] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. 2008.

[14] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882, 2016.

[15] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[16] Hao Liu, Xiangyu Zhu, Zhen Lei, and Stan Z Li. Adaptiveface: Adaptive margin and sampling for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11947–11956, 2019.

[17] Jingtuo Liu, Yafeng Deng, Tao Bai, Zhengping Wei, and Chang Huang. Targeting ultimate accuracy: Face recognition via deep embedding. *arXiv preprint arXiv:1506.07310*, 2015.

[18] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017.

[19] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2537–2546, 2019.

[20] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[21] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild

age database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 51–59, 2017.

[22] Hong-Wei Ng and Stefan Winkler. A data-driven approach to cleaning large face datasets. In *2014 IEEE international conference on image processing (ICIP)*, pages 343–347. IEEE, 2014.

[23] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4004–4012, 2016.

[24] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *bmvc*, volume 1, page 6, 2015.

[25] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[26] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[27] Li Shen, Zhouchen Lin, and Qingming Huang. Relay backpropagation for effective learning of deep convolutional neural networks. In *European conference on computer vision*, pages 467–482. Springer, 2016.

[28] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1891–1898, 2014.

[29] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2892–2900, 2015.

[30] Y Taigman, M Yang, M Ranzato, and L Wolf. Closing the gap to human-level performance in face verification. deepface. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 5, page 6, 2014.

[31] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: l 2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1041–1049. ACM, 2017.

[32] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[33] Mei Wang, Weihong Deng, Jiani Hu, Jianteng Peng, Xunqiang Tao, and Yaohai Huang. Racial faces in-the-wild: Reducing racial bias by deep unsupervised domain adaptation. *arXiv preprint arXiv:1812.00194*, 2018.

[34] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to model the tail. In *Advances in Neural Information Processing Systems*, pages 7029–7039, 2017.

[35] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.

[36] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[37] Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5409–5418, 2017.

[38] Kai Zhao, Jingyi Xu, and Ming-Ming Cheng. Regularface: Deep face recognition via exclusive regularization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1136–1144, 2019.

[39] Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying crosspose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, pages 18–01, 2018.

[40] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.

[41] Q Zhong, C Li, Y Zhang, H Sun, S Yang, D Xie, and S Pu. Towards good practices for recognition & detection. In *CVPR workshops*, volume 1, 2016.

[42] Xiangyu Zhu, Hao Liu, Zhen Lei, Hailin Shi, Fan Yang, Dong Yi, Guojun Qi, and Stan Z Li. Large-scale bisample learning on id versus spot face recognition. *International Journal of Computer Vision*, 127(6-7):684–700, 2019.