Learning Meta Face Recognition in Unseen Domains

Jianzhu Guo[1,2]   Xiangyu Zhu[1,2]   Chenxu Zhao[3]   Dong Cao[1,2]   Zhen Lei[1,2*]   Stan Z. Li[4]

[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]Mininglamp Academy of Sciences, Mininglamp Technology

[4]School of Engineering, Westlake University, Hangzhou, China

{jianzhu.guo, xiangyu.zhu, dong.cao, zlei, szli}@nlpr.ia.ac.cn, zhaochenxu@mininglamp.com

Abstract

Face recognition systems are usually faced with unseen domains in real-world applications and show unsatisfactory performance due to their poor generalization. For example, a well-trained model on webface data cannot deal with the ID vs. Spot task in surveillance scenario. In this paper, we aim to learn a generalized model that can directly handle new unseen domains without any model updating. To this end, we propose a novel face recognition method via meta-learning named Meta Face Recognition (MFR). MFR synthesizes the source/target domain shift with a meta-optimization objective, which requires the model to learn effective representations not only on synthesized source domains but also on synthesized target domains. Specifically, we build domain-shift batches through a domain-level sampling strategy and get back-propagated gradients/meta-gradients on synthesized source/target domains by optimizing multi-domain distributions. The gradients and meta-gradients are further combined to update the model to improve generalization. Besides, we propose two benchmarks for generalized face recognition evaluation. Experiments on our benchmarks validate the generalization of our method compared to several baselines and other state-of-the-arts. The proposed benchmarks and code will be available at https://github.com/cleardusk/MFR.

1. Introduction

Face recognition is a long-standing topic in the research community. Recent works [1, 2, 3, 4, 5, 6, 7, 8] have pushed the performance to a very high level on several common benchmarks, e.g. LFW [9], YTF [10] and MegaFace [11]. These methods are based on the assumption that the training sets like CASIA-Webface [12], MS-Celeb [13] and testing sets have similar distribution. However, in real-world ap-
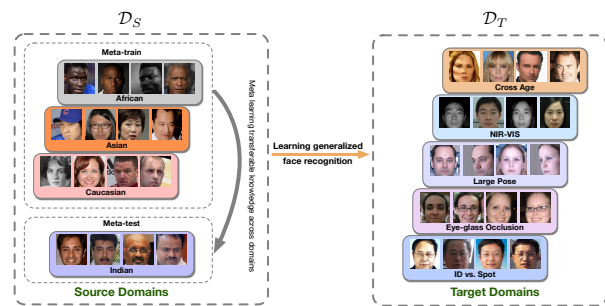

Figure 1: An illustration of our MFR for generalized face recognition problem. The left column contains four source domains of different races, the right includes five target domains: cross-age (CACD-VS), NIR-VIS face matching (CASIA NIR-VIS 2.0), large pose (Multi-PIE), eyeglass occlusion (MeGlass) and ID vs. Spot (Public-IvS), which are unseen in training. By meta-learning on the simulated meta-train/meta-test shifts in source domains, our model learns the transferable knowledge across domains to generalize well on target unseen domains.

plications of face recognition, the model trained on source domains $\mathcal{D}_S$ is usually deployed in another domain $\mathcal{D}_T$ with a different distribution. There are two kinds of scenarios: (i) the target domain $\mathcal{D}_T$ is known and the data is accessible. (ii) the target domain is unseen. Approaches to the first scenario are categorized into domain adaptation for face recognition, where the common setting is that the source domain $\mathcal{D}_S$ contains a labelled face domain and the target domain $\mathcal{D}_T$ is with or without labels. Domain adaption methods try to adapt the knowledge learned from $\mathcal{D}_S$ to $\mathcal{D}_T$ so that the model generalizes well on $\mathcal{D}_T$. The second scenario can be regarded as domain generalization for face recognition, and we call it *Generalized Face Recognition*, which is more common as the trained model is usually deployed in unknown scenarios and faced with unseen data. As illustrated in Fig. 1, the deployed model should be able to generalize to unseen domains without any updating or fine-tuning.

---

*Corresponding author

Compared with domain adaptation, generalized face recognition is less studied and more challenging, since it makes no assumptions about target domains. To the best of our knowledge, there are no relative studies on generalized face recognition problem. A related task is domain generalization on visual recognition, it assumes that the source and target domains share the same label space, and has a small set, e.g., 7 categories [14]. However, generalized face recognition is an open-set problem and has a much larger scale of categories, making existing methods inapplicable.

In this paper, we aim to learn a model for generalized face recognition problem. Once trained on a set of source domains, the model can be directly deployed on an unseen domain without any model updating. Inspired by [14, 15], we propose a novel face recognition framework via meta-learning named Meta Face Recognition (MFR). MFR simulates the source/target domain shift with a meta-optimization objective, which optimizes the model to learn effective face representations not only on synthesized source domains but also on synthesized target domains. Specifically, a domain-level sampling strategy is adopted to simulate the domain shift such that source domains are divided into meta-train/meta-test domains. To optimize multi-domain distributions, we propose three components: 1) the hard-pair attention loss optimizes the local distribution with hard pairs, 2) soft-classification loss considers the global relationship within a batch and 3) domain alignment loss learns to reduce meta-train domains discrepancy by aligning domain centers. These three losses are combined to learn domain-invariant and discriminative face representations. The gradients from meta-train domains and meta-gradients from meta-test domains are finally aggregated by meta-optimization, and are then used to update the network to improve model generalization. Compared with traditional meta-learning methods, our MFR does not need model updating for target domains and can directly handle unseen domains.

Our main contributions include: (i) For the first time, we highlight the generalized face recognition problem, which requires a trained model to generalize well on unseen domains without any updating. (ii) We propose a novel Meta Face Recognition (MFR) framework to solve generalized face recognition, which meta-learns transferable knowledge across domains to improve model generalization. (iii) Two generalized face recognition benchmarks are designed for evaluation. Extensive experiments on the proposed benchmarks validate the efficacy of our method.

2. Related work

Domain Generalization. Domain generalization can be traced back to [16, 17]. DICA [17] adopts the kernel-based optimization to learn domain-invariant features. CCSA [18] can handle both domains adaptation and domain generaliza-

tion problems by aligning a source domain distribution to a target domain distribution. MLDG [14] firstly applies the meta-learning method MAML [15] for domain generalization. Compared with domain adaptation, domain generalization is a less investigated problem. Besides, the above domain generalization works mainly focus on the closed-set category-level recognition problems, where the source and target domains share the same label space. In contrast, our generalized face recognition problem is much more challenging because the target classes are disjoint from the source ones. It means that generalized face recognition is an open-set problem rather than the closed-set problem like MLDG [14], and we must handle the domain gap and the disjoint label space simultaneously. One related work is DIMN [19], but it differs from ours in both task and method.

Meta Learning. Recent meta-learning studies concentrate on: (i) learning a good weight initialization for fast adaptation on a new task, such as the foundational work MAML [15] and its variants Reptile [20], meta-transfer learning [21], iMAML [22] and so on. (ii) learning an embedding space with a well-designed classifier that can directly classify samples on a new task without fast adaptation [23, 24, 25]. (iii) learning to predict the classification parameters [26, 27] after pre-training a good feature extractor on the whole training set. These works focus on few-shot learning, where the common setting is that the target task has very few data points (1/5/20 shots per class). In contrast, generalized face recognition should handle thousands of classes, making it more challenging and generally applicable. Our approach is most related to MAML [15] that tries to learn a transferable weight initialization. However, MAML requires fast adaptation on a target task, while our MFR does not require any model updating as target domains are unseen.

3. Methology

This section describes the proposed MFR to address generalized face recognition problem. MFR consists of three parts: (i) the domain-level sampling strategy. (ii) three losses for optimizing multi-domain distributions to learn domain-invariant and discriminative face representations. (iii) the meta-optimization procedure to improve model generalization, shown in Fig. 3. The overview is shown in Fig. 2 and Algorithm 1.

3.1. Overview

In the training stage, we have access to N source domains $\mathcal{D}_S = \{\mathcal{D}_1^S, \cdots, \mathcal{D}_N^S \mid N > 1\}$, and each domain $\mathcal{D}_i^S = \{(x_j^i, y_j^i)\}$ has its own label set. In the testing phase, the trained model is evaluated on one or several unseen target domains, $\mathcal{D}_T = \{\mathcal{D}_1^T, \cdots, \mathcal{D}_M^T \mid M \geq 1\}$, without any model updating. Besides, the label sets $\mathcal{Y}_T$ of the target domains are disjoint from the label sets $\mathcal{Y}_S$ of source domains,
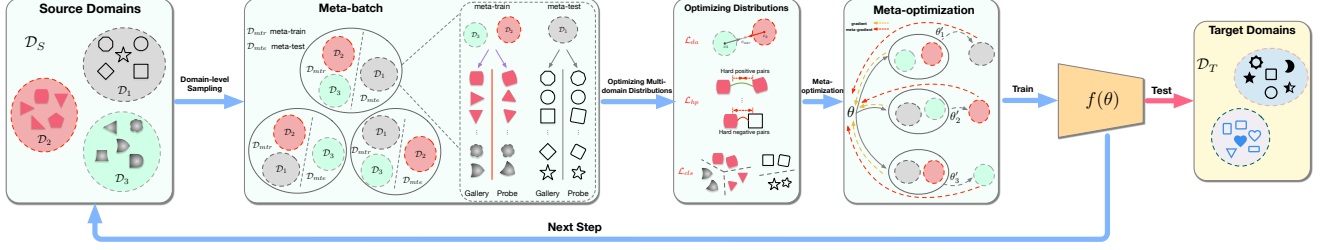
Figure 2: Overview of our proposed MFR. Three source domains are presented in this figure for a demonstration. Each symbol represents a face image for convenience. MFR consists of three parts: domain-level sampling for simulating domain shifts, multi-domain distributions optimization to learn face representations and meta-optimization procedure to improve model generalization. Once trained on source domains, the model can be directly deployed on target unseen domains.

thus making ours problem open-set. During training, we define a single model represented by a parametrized function $f(\theta)$ with parameters θ. Our proposed MFR aims to train θ on source domains $\mathcal{D}_S$, such that it can generalize well on target unseen domains $\mathcal{D}_T$, as illustrated in Fig. 1.

3.2. Domain-level Sampling

To achieve domain generalization, we split source domains into meta-train and meta-test domains during each training iteration. Specifically, we split N source domains $\mathcal{D}_S$ into $N-1$ domains $\mathcal{D}_{mtr}$ for meta-train and 1 target domain $\mathcal{D}_{mte}$ for meta-test, simulating the domain shift problem existed when deployed in real-world scenarios. In this way, the model is encouraged to learn transferable knowledge about how to generalize well on the unseen domains with different distributions. We further build a meta-batch consisting of several batches as follows: (i) we iterate on N source domains; (ii) in the i-th iteration, $\mathcal{D}_i^S$ is selected as the meta-test domain $\mathcal{D}_{mte}$; (iii) the rest ones as meta-train domains $\mathcal{D}_{mtr}$; (iv) we randomly choose B identities in meta-train domains and B identities in meta-test domain, and two face images are selected for each identity, in which one as the gallery the other one as probe. Therefore, a meta-batch of N batches is built. Our model is then updated by the accumulated gradients of each meta-batch. The details are illustrated in Algorithm 1. Different from MAML [15], our sampling is domain-level for open-set face recognition. MLDG [14] also performs a similar sampling, but their domains are randomly divided in each training iteration and no meta-batch is built.

3.3. Optimizing Multi-domain Distributions

To aggregate back-propagated gradients within each batch, we optimize multi-domain distributions such that the same identities are mapped into nearby representation and different identities are mapped apart from each other. Traditional metric losses like contrastive [28, 29] and triplet [3] take randomly sampled pairs or triplets to build the training batches. These batches consist of lots of easy pairs or triplets, leading to the slow convergence of training. To alle-

viate it, we propose to optimize and learn domain-invariant and discriminative representations with three components. The hard-pair attention loss optimizes the local distribution with hard pairs, the soft-classification loss considers the global distribution within a batch and the domain alignment loss learns to align domain centers.

Hard-pair Attention Loss. Hard-pair attention loss focuses on optimizing hard positive and negative pairs. A batch of B identities are sampled and each identity contains a gallery face and a probe face. We denote the input as $\mathcal{X}$, the gallery and probe embeddings are extracted: $F_g = f(\mathcal{X}_g; \theta) \in R^{B \times C}$, $F_p = f(\mathcal{X}_p; \theta) \in R^{B \times C}$, where C is the dimension length. After l_2 normalization on F_g and F_p, we can efficiently construct a similarity matrix by computing $M = F_g F_p^T \in R^{B \times B}$. Then we use a positive threshold τ_p and negative threshold τ_n to filter out the hard positive pairs and negative pairs: $\mathcal{P} = \{i | M_{i,i} < \tau_p\}$ and $\mathcal{N} = \{(i,j) | M_{i,j} > \tau_p, i \neq j\}$. This operation just needs $O(B^2 \log(B))$ complexity and it can be formulated as:

$$\mathcal{L}_{hp} = \frac{1}{2|\mathcal{P}|} \sum_{i \in \mathcal{P}} \|F_{g_i} - F_{p_i}\|_2^2 - \frac{1}{2|\mathcal{N}|} \sum_{(i,j) \in \mathcal{N}} \|F_{g_i} - F_{p_j}\|_2^2,$$

$$(1)$$

where $\mathcal{P}$ is indices of hard positive pairs determined by τ_p, $\mathcal{N}$ is indices of hard negative pairs determined by τ_n.

Soft-classification Loss. Hard-pair attention loss only concentrates on hard pairs and tends to converge to a local optimum. To alleviate it, we introduce a specific soft-classification loss to perform classification within a batch. The loss is formulated as:

$$\mathcal{L}_{cls} = \frac{1}{2B} \sum_{i=1}^{B} \Big(\text{CE}(y_i, s \cdot F_{g_i} W^T) + \text{CE}(y_i, s \cdot F_{p_i} W^T) \Big),$$

$$(2)$$

where $y_i = i$ indicates the i-th identity, $F_{g_i} W^T$ or $F_{p_i} W^T$ is the logit of i-th identity and s is a fixed scaling factor. W is initialized as $(F_g + F_p)/2 \in R^{B \times C}$ and each row of W is l_2 normalized.
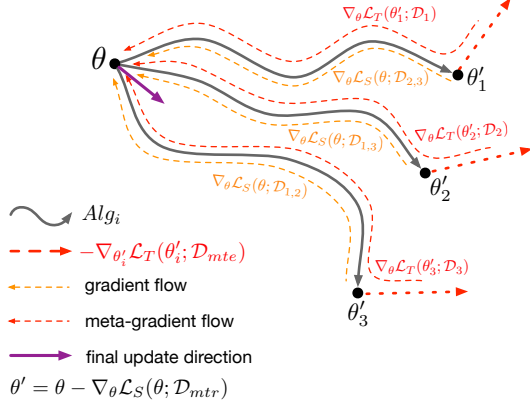
Figure 3: Overview of the meta-optimization procedure in a meta-batch. Given three source domains $\mathcal{D}_{1,2,3}$, a meta batch contains three meta-train/meta-test divisions: $\mathcal{D}_{2,3}$ / $\mathcal{D}_1$, $\mathcal{D}_{1,3}$ / $\mathcal{D}_2$, $\mathcal{D}_{1,2}$ / $\mathcal{D}_3$. Each division contributes a gradient from meta-train and a meta-gradient from meta-test. The model is finally updated towards a direction that performs well on both meta-train and meta-test domains by accumulating all the gradients and meta-gradients.

Domain Alignment Loss. We find negative pairs across meta-train domains tend to be easier than within domains. By adding a domain alignment regularization to make the embeddings domain-invariant, we can reduce domain gap of different meta-train domains. Besides, negative pairs across meta-train domains become harder, which is beneficial to learn more discriminative representations. To perform domain alignment, we make the mean embeddings of multiple meta-train domains close to each other. Specifically, we first calculate the embedding center of all mean embeddings of meta-train domains, then optimize the discrepancies between all mean embeddings and this embedding center. The domain alignment loss is only applied on meta-train domains, as meta-test has only one domain. The loss is formulated as:

$$c_j = \frac{1}{B} \sum_{i=1}^{B} \left((F_{g_i}^{\mathcal{D}_j} + F_{p_i}^{\mathcal{D}_j})/2 \right),$$

$$c_{mtr} = \frac{1}{n} \sum_{j=1}^{n} c_j, \qquad (3)$$

$$\mathcal{L}_{da} = \frac{1}{n} \sum_{j=1}^{n} \| s \cdot (c_j - c_{mtr}) \|_2^2,$$

where F_{g_i}, F_{p_i} are normalized embeddings, c_j is the mean embedding within a batch sampled from domain $\mathcal{D}_j$, c_{mtr} is the embedding center of all mean embeddings of meta-train domains, n is the number of meta-train domains and s is the scaling factor. In meta-optimization, we will adaptively utilize the back-propagated signals from these three losses to improve the model generalization.

3.4. Meta-optimization

This section describes how the model is optimized to improve model generalization. The whole meta-optimization procedure is summarized in Algorithm 1 and illustrated in Fig. 3.

Meta-train. Based on domain-level sampling, during each batch within a meta-batch, we sample $N-1$ source domains $\mathcal{D}_{mtr}$ and B pairs of images $\mathcal{X}_S$ from $\mathcal{D}_{mtr}$. Then we conduct the proposed losses in each batch as follows:

$$\mathcal{L}_S = \mathcal{L}_{hp}(\mathcal{X}_S; \theta) + \mathcal{L}_{cls}(\mathcal{X}_S; \theta) + \mathcal{L}_{da}(\mathcal{X}_S; \theta), \quad (4)$$

where θ represents the model parameters. The model is next updated by gradient ∇_θ as: $\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_S(\theta)$. This update step is similar to the conventional metric learning.

Meta-test. In each batch, the model is also tested on the meta-test domain $\mathcal{D}_{mte}$. This testing procedure simulates the evaluating on an unseen domain with a different distribution, so as to make the the model to learn to generalize across domains. We also sample B pairs of images $\mathcal{X}_T$ from the meta-test domain $\mathcal{D}_{mte}$. Then the loss is conducted on the updated parameters θ' as below:

$$\mathcal{L}_T = \mathcal{L}_{hp}(\mathcal{X}_T; \theta') + \mathcal{L}_{cls}(\mathcal{X}_T; \theta'). \quad (5)$$

Summary. To optimize the meta-train and meta-test simultaneously, the final MFR objective is:

$$\arg\min_\theta \gamma \mathcal{L}_S(\theta) + (1 - \gamma)\mathcal{L}_T(\theta - \alpha \mathcal{L}'_S(\theta)), \quad (6)$$

where α is the meta-train step-size and γ balances meta-train and meta-test. This objective can be understood as: *optimize the model parameters, such that after updating on the meta-train domains, the model also performs well on the meta-test domain.* From another perspective, the second term of Eqn. 6 serves as an extra regularization to update the model with high order gradients, and we call it meta-gradients. For example, given three source domains $\mathcal{D}_S = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$, a meta-batch consists of three meta-train/meta-test divisions: $\mathcal{D}_2, \mathcal{D}_3/\mathcal{D}_1$, $\mathcal{D}_1, \mathcal{D}_3/\mathcal{D}_2$ and $\mathcal{D}_1, \mathcal{D}_2/\mathcal{D}_3$. For each division or batch, a gradient and a meta-gradient are back-propagated on meta-train and meta-test, respectively. By accumulating all the gradients and meta-gradients in the meta-batch, the model is finally optimized to perform well on both meta-train and meta-test domains. Fig. 3 illustrates how the gradients and meta-gradients flow on the computation graph.

4. Experiments

To evaluate our proposed MFR for generalized face recognition problem, we conduct several experiments on two proposed benchmarks.

Algorithm 1: MFR for generalized face recognition problem.

Input: Source (training) domains
$$\mathcal{D}_S = \{\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_N\}.$$
Init: A pre-trained model $f(\theta)$ parametrized by θ, hyperparameters α, β, γ and batch-size of B.

---

1 **for** *ite in max_iterations* **do**
2   Init the gradient g_θ as $\mathbf{0}$;
    // For a meta-batch
3   **for** *each $\mathcal{D}_{mte}$ in $\mathcal{D}_S$* **do**
       // For a batch
4   Sampling remaining domains as $\mathcal{D}_{mtr}$;
5   **Meta-train:**
6   Sampling B paired images $\mathcal{X}_S$ from B identities of meta-train domains $\mathcal{D}_{mtr}$;
7        $\mathcal{L}_S = \mathcal{L}_{hp}(\mathcal{X}_S;\theta) + \mathcal{L}_{cls}(\mathcal{X}_S;\theta) + \mathcal{L}_{da}(\mathcal{X}_S;\theta)$;
8   Update model parameters by:
         $\theta' = \theta - \alpha\nabla_\theta\mathcal{L}_S(\theta)$;
9   **Meta-test:**
10   Sampling B paired images $\mathcal{X}_T$ from B identities of the meta-test domain $\mathcal{D}_{mte}$;
11       $\mathcal{L}_T = \mathcal{L}_{hp}(\mathcal{X}_T;\theta') + \mathcal{L}_{cls}(\mathcal{X}_T;\theta')$;
12   **Gradient aggregation:**
13       $g_\theta \leftarrow g_\theta + \gamma\nabla_\theta\mathcal{L}_S + (1-\gamma)\nabla_\theta\mathcal{L}_T$;
14   **end**
15   **Meta-optimization:**
16   Update $\theta \leftarrow \theta - \frac{\beta}{N}g_\theta$ by SGD;
17 **end**

4.1. GFR Benchmark and Protocols

Generalized face recognition has not attracted much attention and we do not have a common protocol for evaluation, thus we introduce two well-designed benchmarks to evaluate the generalization of a model. One benchmark is for crossing race evaluation named GFR-R and another one is crossing facial variety named GFR-V. We use variety here to emphasize that there is a large gap between source domains and target unseen domains on GFR-V.

In a real-world scenario, a large-scale base dataset like MS-Celeb [13] is usually available for pre-training, but the model may generalize poorly on a new domain with a different distribution. To simulate it, we use MS-Celeb as the base dataset. RFW [35] is originally proposed to study the racial bias in face recognition and it labels four racial datasets (Caucasian, Asian, African, Indian) from MS-Celeb. We choose to select these four datasets as our four racial domains. Note that RFW [35] overlaps MS-Celeb [13], we remove all the overlapped identities from MS-Celeb according to the identity keyword, thus building

our base dataset named MS-Celeb-NR[1], which means MS-Celeb without RFW. MS-Celeb-NR can be regarded as an independent base dataset of four racial ones.

GFR-R. Each race has about 2K~3K identities. We randomly choose 1K identities for testing and the remaining 1K~2K identities for training. The dataset details are shown in Table 1. In our experiment setting, each race is regarded as one domain. We randomly select three domains in four as source domains and and the rest one as the testing domain, which is not accessible in training. Therefore, we build four sub-protocols for GFR-R, shown in Table 2.

GFR-V. The GFR-V benchmark is for crossing facial variety evaluation, which is a harder setting and can better reflect the generalization ability of a model. As is shown in Table 2, four racial datasets (Caucasian, Asian, African, Indian) are treated as source domains, and five datasets are as target domains. Specifically, the target datasets include CACD-VS [30], CASIA NIR-VIS 2.0 [31], MultiPIE [32], MeGlass [33], Public-IvS [34]. For CASIA NIR-VIS 2.0, we follow the standard protocol in View 2 evaluation [31] and we report the average value of 10 folds. For MeGlass and Public-IvS, we follow the standard testing protocols [34, 33]. For CACD-VS, in addition to the standard protocol [30], we use the provided 2,000 positive cross-age image pairs and split them into gallery and probe for our ROC/Rank-1 evaluation. For Multi-PIE, we select 337 identities and each identity contains about 3~4 frontal gallery images and 3~4 probe images with the 45° view.

Benchmark Protocols. For each image, the features from both the original image and the flipped one are extracted then concatenated as the final representation. The score is measured by the cosine distance of two representations. For performance evaluation, we use the receiver operating characteristic (ROC) curve and Rank-1 accuracy. For ROC, we report the verification rate (VR) at low false acceptance rate (FAR) like 1%, 0.1% and 0.01%. For Rank-1 evaluation, each probe image is matched to all gallery images, if the top-1 result is within the same identity, it is correct.

4.2. Implementation Details

Our experiments are based on PyTorch [37]. The random seed is set to a fixed value 2019 in comparative experiments for fair comparisons. We use a 28-layer ResNet as our backbone, but with a channel-number multiplier of 0.5. Our backbone has only 128.7M FLOPs and 4.64M parameters, which is relatively light-weighted. The dimension of the output embedding is 256. The model is pre-trained on MS-Celeb-NR with CosFace [38]. During training, all faces are cropped and resized to 120×120. The inputs are then normalized by subtracting 127.5 and being divided by 128. The meta-train step-size α, the meta optimization step-size

---

[1]We will release the list of MS-Celeb-NR.

Facial Variety	Dataset	#Train IDs	#Train images	#Test IDs		#Test images	
				#Gallery IDs	#Probe IDs	#Gallery images	#Probe images
Race	Caucasian	1,957	6,757	1,000	1,000	1,000	1,000
Race	Asian	1,492	5,784	1,000	1,000	1,000	1,000
Race	African	1,995	6,938	1,000	1,000	1,000	1,000
Race	Indian	1,984	6,857	1,000	1,000	1,000	1,000
Age	CACD-VS [30]	-	-	2,000	2,000	2,000	2,000
Illumination	CASIA NIR-VIS 2.0 [31]	-	-	358	363	358	6,208
Pose	MultiPIE [32]	-	-	337	337	1,184	1,181
Occlusion	MeGlass [33]	-	-	1,710	1,710	3,420	3,420
Heterogeneity	Public-IvS [34]	-	-	1,262	1,262	1,262	4,241

Table 1: The statistics of all involved datasets. CASIA NIR-VIS 2.0 has 10 folds and the first fold is shown. The other folds own similar statistics.

Protocol		Source Domains	Target Domain(s)
GFR-R	I	Caucasian Asian African	Indian
	II	Caucasian Asian Indian	African
	III	Caucasian African Indian	Asian
	IV	Asian African Indian	Caucasian
GFR-V		Caucasian Asian African Indian	CACD-VS CASIA NIR-VIS 2.0 MultiPIE MeGlass Public-IvS

Table 2: The GFR-R and GFR-V benchmarks. Source domains are for training, target domains are for evaluation and are unseen during training.

Protocol	Method	VR (%)			Rank-1 (%)
		FAR=1%	FAR=0.1%	FAR=0.01%	
GFR-R I (Indian)	Base	94	82.2	64.65	80.3
	Base-Agg	94.1	80.9	65.3	81
	Base-FT rnd.	62.5	39	21.05	39.3
	Base-FT imp. [36]	87	69.9	51.2	69.6
	MLDG [14]	94.2	83	66.3	80.5
	MFR (Ours)	**95.4**	**86.1**	**71.4**	**83.1**
GFR-R II (African)	Base	91.6	74.5	55.4	73.1
	Base-Agg	90.5	74.8	56.3	74
	Base-FT rnd.	26.2	10.9	3.5	21
	Base-FT imp. [36]	78.7	56.6	36.45	57.9
	MLDG [14]	91.9	74.8	55.7	73.8
	MFR (Ours)	**92.3**	**79.4**	**60.8**	**75.2**
GFR-R III (Asian)	Base	91.89	77.98	60.86	75.98
	Base-Agg	91.49	78.08	59.41	76.28
	Base-FT rnd.	40.44	17.32	7.67	27.53
	Base-FT imp. [36]	80.58	57.56	39.79	61.86
	MLDG [14]	92.29	78.28	60.3	76.68
	MFR (Ours)	**93.49**	**80.7**	**62.56**	**78.68**
GFR-R IV (Caucasian)	Base	96.6	89.6	78.6	86.6
	Base-Agg	97	88.1	79.1	86.8
	Base-FT rnd.	61.1	36.2	18.9	36.7
	Base-FT imp. [36]	91.5	78.2	63.4	76.8
	MLDG [14]	96.8	89.6	79.15	86.3
	MFR (Ours)	**98.2**	**92.9**	**81.1**	**88.9**

Table 3: Comparative results of the GFR-R benchmark. rnd. means random initializing the classification weight template, imp. is weight-imprinted.

β, the weight γ balancing meta-train and meta-test loss are initialized to 0.0004, 0.0004, 0.5, respectively. Batch-size B is set to 128 and the scaling factor s of both the soft-classification loss and domain alignment loss are set to 64. The step-size α and β are decayed with every 1K steps and the decay rate is 0.5. The positive threshold τ_p and negative threshold τ_n are initialized to 0.3, 0.04 and are updated as $\tau_p = 0.3 + 0.1n$ and $\tau_n = 0.04/0.5^n$, where n is the decayed number. For meta-optimization, we use SGD to optimize the network with the weight decay of 0.0005 and momentum of 0.9.

4.3. GFR-R Comparisons

Settings. We compare our model with several baselines, including the base model and several domain aggregation baselines. To further compare our method with other domain generalization methods, we adapt MLDG [14] to an open-set setting, so that it can be applied in our protocols. The results are shown in Table 3. For four protocols in GFR-R, we report the VRs at low FAR 1%, 0.1%, 0.01%, and the Rank-1 accuracy. Specifically, our comparisons include: (i) *Base*: the model pre-trained only on MS-Celeb-NR using CosFace [38]. Note that MS-Celeb-NR has no overlapped identities with four racial datasets (Caucasian, Asian, African and Indian) and can be considered as an independent dataset. (ii) *Base-Agg*: the model trained on MS-Celeb-NR and the aggregation of source domains using CosFace [38]. Take GFR-R-I as an example, *Base-Agg* is trained on MS-Celeb-NR and three source domains Caucasian, Asian, African jointly. This is for the fair comparison with our MFR, where the same training datasets are involved. (iii) *Base-FT rnd.*: the base model further fine-tuned on the aggregation of source domains. The classification template of the last FC-layer is randomly initialized. (iv) *Base-FT imp.*: the base model further fine-tuned on the aggregation of source domains, but the classification template is initialized as the mean of embeddings of the corresponding identities. It is refined from weight-imprinted [36]. (v) *MLDG*: MLDG [14] adapted for generalized face recognition problem.

Results. From the results in Table 3, the following observations can be made: (i) Overall, our method achieves the

best result on four GFR-R protocols among all compared settings and methods. (ii) The base model pre-trained on MS-Celeb-NR is strong, but not generalizes well for target domains, especially for Indian, African, Asian. The reason may be that MS-Celeb-NR is occupied by Caucasian people. (iii) Jointly training on MS-Celeb-NR and source domains performs slightly better than the base model, but is still not comparable to our MFR method. (iv) The performance of *Base-FT rnd.* declines dramatically and we attribute it to over-fitting on source domains. Weight-imprinted (*Base-FT imp.*) can reduce such over-fitting to some degree, but its performance is still lower than the base model. (v) MLDG [14], which is originally designed for closed-set and category-level recognition problems, fail to compete with our method on the open-set generalized face recognition problem.

GFR-V (CACD-VS)	VR (%)		Rank-1 (%)	Acc.	AUC.
	FAR=0.01%	FAR=0.001%			
Base	96.55	92.55	96.85	99.35	99.53
Base-Agg	96.75	92.98	97.15	99.42	99.6
MLDG [14]	96.75	92.9	97.25	99.45	99.57
LF-CNNs [39]	-	-	-	98.5	99.3
Human, Voting [40]	-	-	-	94.2	99
OE-CNNs [41]	-	-	-	99.2	99.5
AIM+CAFR [42]	-	-	-	99.76	-
MFR (Ours)	**97.25**	**94.05**	**97.8**	**99.78**	**99.81**

Table 4: Comparative results on CACD-VS.

GFR-V CASIA NIR-VIS 2.0	VR (%)			Rank-1 (%)
	FAR=1%	FAR=0.1%	FAR=0.01%	
Base	97.8	89.89	69.27	93.18
Base-Agg	98.31	90.47	**71.35**	94.29
MLDG [14]	98.28	90.44	69.32	93.56
IDR [43]	98.9	95.7	-	**97.3**
WCNN [44]	**99.4**	**97.6**	-	**98.4**
MFR (Ours)	**99.32**	**95.97**	**81.92**	96.92

Table 5: Comparative results on CASIA NIR-VIS 2.0. The highest two results are highlighted.

GFR-V (Multi-PIE)	VR (%)			Rank-1 (%)
	FAR=0.1%	FAR=0.01%	FAR=0.001%	
Base	99.92	98.83	61.54	99.75
Base-Agg	99.92	98.96	68.49	99.82
MLDG [14]	99.84	98.87	62.95	99.83
MFR (Ours)	**100**	**99.96**	**74.54**	**99.92**

Table 6: Comparative results on MultiPIE.

GFR-V (MeGlass)	VR (%)			Rank-1 (%)
	FAR=0.01%	FAR=0.001%	FAR=0.0001%	
Base	85.92	71.96	53.5	97.6
Base-Agg	86.77	73.5	54.96	97.69
MLDG [14]	85.54	69.23	49.32	**97.81**
Face Syn. [33]	**90.14**	**80.32**	**66.92**	96.73
MFR (Ours)	**90.79**	**80.86**	**66.15**	**98.57**

Table 7: Comparative results on MeGlass. The highest two results are highlighted.

GFR-V (Public-IvS)	VR (%)			Rank-1 (%)
	FAR=0.1%	FAR=0.01%	FAR=0.001%	
Base	94.38	86.71	74.83	92.74
Base-Agg	94.24	87.1	74.5	92.85
MLDG [14]	94.96	87.35	75.54	**93.3**
Contrastive [29]	96.52	91.71	84.54	-
LBL [34]	**98.83**	**97.21**	**93.62**	-
MFR (Ours)	**96.66**	**92.96**	**85.28**	**95.82**

Table 8: Comparative results on Public-IvS. The highest two results are highlighted.

Method	Base	Base-Agg	MLDG [14]	**MFR (Ours)**
LFW	99.57	99.60	99.43	**99.77**

Table 9: Comparative results on LFW.

4.4. GFR-V Comparisons

The GFR-V benchmark is for crossing facial variety evaluation, which can better reflect model generalization.

Settings. We compare our model with two strong baselines *Base*, *Base-Agg*, an adapted MLDG [14] and other competitors if existed. Since the standard protocols differ among five target domains, we show them separately in Table 4, 5, 6, 7, 8.

CACD-VS. CACD-VS [30] is for cross-age evaluation, where each pair of images contain a young face and an old one. We report ROC/Rank-1 as well as the standard protocol provided. Other competitors are only evaluated on the standard protocol. The results in Table 4 show that our MFR not only beats the baselines but also the competitors, which use cross-age datasets for training.

CASIA NIR-VIS 2.0. In CASIA NIR-VIS 2.0 [31], gallery images are collected under visible lighting, while the probe one is under near infrared lighting, thus the modality gap is huge. Table 5 shows: (i) we achieve great performance improvements from 89.89% (69.27%) of *Base* to 95.97% (81.92%) when FAR=0.1% (0.01%). (ii) even with such a huge modality gap, our performance is comparable to several CNN-based methods [43, 44], which use MS-Celeb for pre-training and the target domain NIR-VIS dataset for fine-tuning. In comparison, our model has not seen any NIR samples during training.

Multi-PIE. We compare our model with two baselines and MLDG for cross-pose evaluation using Multi-PIE. Table 6 validates the improvements of our MFR over baselines and MLDG.

MeGlass. MeGlass [33] focuses on the effect of eyeglass occlusion for face recognition. We select the hardest IV protocol for evaluation. As shown in Table 7, our method promotes the performance from 71.96% (53.5%) on *Base* to 80.86% (66.15%) when at a low FAR 0.001%, which is even slightly better than [33], which synthesizes wearing-eyeglass image for the whole MS-Celeb for training.

Public-IvS. Public-IvS [34] is a testbed for ID vs. Spot (IvS) verification. Compared to *Base* and *Base-Agg*, our method greatly improves the generalization performance.

The other two competitors are all pre-trained on MS-Celeb and fine-tuned on CASIA-IvS, which has more than 2 million identities and each identity has one ID and Spot face. Even so, our method still performs slightly better than Contrastive [29].

LFW. We perform an extensive evaluation on LFW [9], shown in Table. 9. The results demonstrate that our method also generalizes better than baselines on a similar target domain.

The above results show that our method achieves great improvement than baselines, and the performance is competitive to the best supervised / non-generalization methods. For a real-world face recognition application, our method is the first choice because it generalizes well on all target domains with competitive performances.

Protocol	Method	VR (%)			Rank-1 (%)
		FAR=1%	FAR=0.1%	FAR=0.01%	
GFR-R I (Indian)	w/o hp.	95.1	84.5	69.2	82.2
	w/o cls.	95.3	84.3	69	82.3
	w/o da.	95.2	84.9	70.8	82.7
	w/o meta ($\alpha = 0$)	94.8	84.3	68.35	81.1
	first order	95.3	85.7	70.9	82.6
	Ours-full	**95.4**	**86.1**	**71.4**	**83.1**
GFR-R II (African)	w/o hp.	92	77.9	59.2	74.6
	w/o cls.	92.1	78.9	59.3	74.6
	w/o da.	92.1	78.6	59.4	74.8
	w/o meta ($\alpha = 0$)	91.9	77.6	57.75	74.8
	first order	92	78.05	59.5	75.2
	Ours-full	**92.3**	**79.4**	**60.8**	**75.2**
GFR-R III (Asian)	w/o hp.	93.39	79.9	61.56	77.78
	w/o cls.	93.29	80.4	62.1	78.08
	w/o da.	93.49	79.68	61.76	77.78
	w/o meta ($\alpha = 0$)	92.89	79.2	60.7	77.28
	first order	**93.49**	79.9	61.7	77.88
	Ours-full	**93.49**	**80.7**	**62.56**	**78.68**
GFR-R IV (Caucasian)	w/o hp.	98.2	91.3	80.4	87.8
	w/o cls.	98.3	92.6	80.4	88.4
	w/o da.	**98.4**	92.4	80.5	87.5
	w/o meta ($\alpha = 0$)	97.3	91.1	79.6	87.3
	first order	97.9	91.8	80.1	87.7
	Ours-full	98.2	**92.9**	**81.1**	**88.9**

Table 10: Ablative results of the GFR-R benchmark. hp. is the hard-pair attention loss, cls. is the soft-classification loss and da. is the domain alignment loss on meta-train domains.

4.5. Ablation Study and Analysis

Contribution of Different Components. To evaluate the contributions of different components, we compare our full MFR with four degraded versions. The first three components are the hard-pair attention loss, soft-classification loss and domain alignment loss, which are designed for learning domain-invariant and discriminative representations. The fourth component is the meta-gradient. If α is set to 0 in Eqn. 6, the objective is degraded to the sum of meta-train and meta-test and there is no meta-gradient computation. Table 10 shows that each component contributes to the performance. Among three components, the meta-gradient is the most important one. For example, in GFR-R I, the performance drops from 71.4% to 68.35% when FAR=0.01% without the meta-gradient.

First Order Approximation. The meta-gradient needs high order derivatives and is computationally expensive. Therefore, we compare it with the first order approximation.
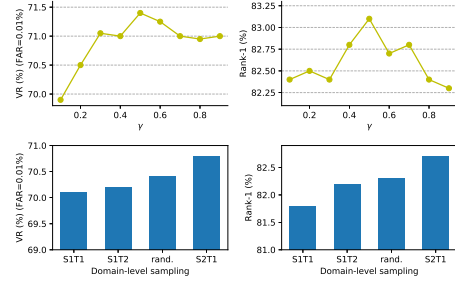

Figure 4: Ablation results on GFR-R I (Indian) protocol, with different γ and domain-level sampling strategies.

To achieve the first order approximation, we only need to change $(1-\gamma)\nabla_\theta \mathcal{L}_T$ in Algorithm 1 to $(1-\gamma)\nabla_{\theta'}\mathcal{L}_T$ in the gradient aggregation step. From Table 10, we can see that the performance of the first order approximation is close to high order. Considering that the first order approximation takes only about 82% GPU memory and 63% time (in our setting) of the high order, the first order approximation is a practical substitute for the high order implementation.

Impact of γ. In Eqn. 6, γ is a hyperparameter weighting the meta-train and meta-test losses. The ablative results are shown in Fig. 4. A proper value 0.5 gives the best result, which indicates the meta-train and meta-test domains should be equally learned.

Domains-level Sampling. Since domain alignment loss cannot be applied when there is only one domain in meta-train, we remove it for fair comparisons. For each batch, SmTn ($m, n \in \{(1,1),(1,2),(2,1)\}$) means sampling m domain as meta-train and another n as meta-test. rand. means randomly choosing m domains as meta-train (m is a random number) and remaining one as meta-test. Fig. 4 shows that the setting $m = 2$ and $n = 1$ performs best.

5. Conclusion

In this paper, we highlight generalized face recognition problem and propose a Meta Face Recognition (MFR) method to address it. Once trained on a set of source domains, the model can be directly deployed on target domains without any model update. Extensive experiments on two newly defined generalized face recognition benchmarks validate the effectiveness of our proposed MFR. We believe generalized face recognition problem is of great importance for practical applications, and our work is an important avenue for future works.

Acknowledgments

References

[1] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014. 1

[2] 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. 1

[3] 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. 1, 3

[4] 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. 1

[5] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, 2019. 1

[6] Xiaobo Wang, Shuo Wang, Shifeng Zhang, Tianyu Fu, Hailin Shi, and Tao Mei. Support vector guided softmax loss for face recognition. *arXiv preprint arXiv:1812.11317*, 2018. 1

[7] Xiaobo Wang, Shuo Wang, Jun Wang, Hailin Shi, and Tao Mei. Co-mining: Deep face recognition with noisy labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9358–9367, 2019. 1

[8] Xiaobo Wang, Shifeng Zhang, Shuo Wang, Tianyu Fu, Hailin Shi, and Tao Mei. Mis-classified vector guided softmax loss for face recognition. *arXiv preprint arXiv:1912.00833*, 2019. 1

[9] 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. 1, 8

[10] Lior Wolf, Tal Hassner, and Itay Maoz. *Face recognition in unconstrained videos with matched background similarity*. IEEE, 2011. 1

[11] 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. 1

[12] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014. 1

[13] 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. 1, 5

[14] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 2, 3, 6, 7

[15] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017. 2, 3

[16] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A Efros, and Antonio Torralba. Undoing the damage of dataset bias. In *European Conference on Computer Vision*, pages 158–171. Springer, 2012. 2

[17] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *International Conference on Machine Learning*, pages 10–18, 2013. 2

[18] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5715–5725, 2017. 2

[19] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 719–728, 2019. 2

[20] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 2

[21] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2019. 2

[22] Aravind Rajeswaran, Chelsea Finn, Sham Kakade, and Sergey Levine. Meta-learning with implicit gradients. *arXiv preprint arXiv:1909.04630*, 2019. 2

[23] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in neural information processing systems*, pages 3630–3638, 2016. 2

[24] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017. 2

[25] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018. 2

[26] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L Yuille. Few-shot image recognition by predicting parameters from activations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7229–7238, 2018. 2

[27] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4367–4375, 2018. 2

[28] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006. 3

[29] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems*, pages 1988–1996, 2014. 3, 7, 8

[30] 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. 5, 6, 7

[31] Stan Li, Dong Yi, Zhen Lei, and Shengcai Liao. The casia nir-vis 2.0 face database. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 348–353, 2013. 5, 6, 7

[32] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010. 5, 6

[33] 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. 5, 6, 7

[34] 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. 5, 6, 7

[35] 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. 5

[36] Hang Qi, Matthew Brown, and David G Lowe. Low-shot learning with imprinted weights. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5822–5830, 2018. 6

[37] 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. In *NIPS Autodiff Workshop*, 2017. 5

[38] 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. 5, 6

[39] Yandong Wen, Zhifeng Li, and Yu Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4893–4901, 2016. 7

[40] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 17(6):804–815, 2015. 7

[41] Yitong Wang, Dihong Gong, Zheng Zhou, Xing Ji, Hao Wang, Zhifeng Li, Wei Liu, and Tong Zhang. Orthogonal deep features decomposition for age-invariant face recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 738–753, 2018. 7

[42] Jian Zhao, Yu Cheng, Yi Cheng, Yang Yang, Fang Zhao, Jianshu Li, Hengzhu Liu, Shuicheng Yan, and Jiashi Feng. Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9251–9258, 2019. 7

[43] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan. Learning invariant deep representation for nir-vis face recognition. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. 7

[44] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan. Wasserstein cnn: Learning invariant features for nir-vis face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1761–1773, 2018. 7