# BioNet: A Biologically-inspired Network for Face Recognition

Pengyu Li
Terminus Labs, Terminus Group, China.
lipengyu007@gmail.com

## Abstract

*Recently, whether and how cutting-edge Neuroscience findings can inspire Artificial Intelligence (AI) confuse both communities and draw much discussion. As one of the most critical fields in AI, Computer Vision (CV) also pays much attention to the discussion. To show our ideas and experimental evidence to the discussion, we focus on one of the most broadly researched topics both in Neuroscience and CV fields,* i.e.*, Face Recognition (FR). Neuroscience studies show that face attributes are essential to the human face-recognizing system. How the attributes contribute also be explained by the Neuroscience community. Even though a few CV works improved the FR performance with attribute enhancement, none of them are inspired by the human face-recognizing mechanism nor boosted performance significantly. To show our idea experimentally, we model the biological characteristics of the human face-recognizing system with classical Convolutional Neural Network Operators (CNN Ops) purposely. We name the proposed* ***Bio****logically-inspired* ***Net****work as* ***BioNet****. Our BioNet consists of two cascade sub-networks,* i.e.*, the Visual Cortex Network (VCN) and the Inferotemporal Cortex Network (ICN). The VCN is modeled with a classical CNN backbone. The proposed ICN comprises three biologically-inspired modules,* i.e.*, the Cortex Functional Compartmentalization, the Compartment Response Transform, and the Response Intensity Modulation. The experiments prove that: 1) The cutting-edge findings about the human face-recognizing system can further boost the CNN-based FR network. 2) With the biological mechanism, both identity-related attributes (*e.g.*, gender) and identity-unrelated attributes (*e.g.*, expression) can benefit the deep FR models. Surprisingly, the identity-unrelated ones contribute even more than the identity-related ones. 3) The proposed BioNet significantly boosts state-of-the-art on standard FR benchmark datasets. For example, BioNet boosts IJB-B@1e-6 from 52.12% to 68.28% and MegaFace from 98.74% to 99.19%. The source code is released in*[1].

[1]https://github.com/pengyuLPY/BioNet.git

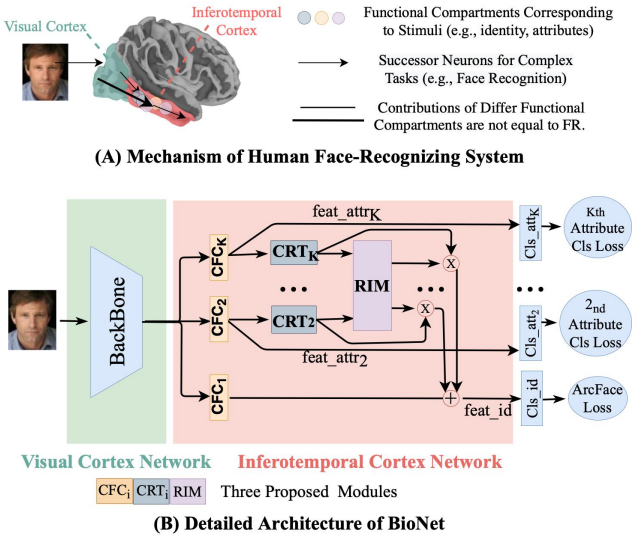


Figure 1. (A) Face recognition mechanism of human brains. (B) Architecture of BioNet.

## 1. Introduction

It sparked much discussion in both communities that Zador, Bengio *et al*. [54] claimed fundamental Neuroscience research must be invested to accelerate AI progress. Some researchers agree with Zador's opinion. For example, LeCun *et al*. [52] and Goodfellow *et al*. [10] proposed the Convolutional Neural Network (CNN) inspired by past classical Neuroscience discoveries about the human visual cortex. However, other researchers have some concerns, *e.g*., 1) Except for the high-level and abstract senses from Neuroscience, can their specific studies support the AI fields? 2) Since CNN [9,10,26,52] was proposed many years ago, few AI works have been inspired by recent Neuroscience findings. It results in a lack of evidence to support that the latest Neuroscience studies can continue to drive AI progress.

To show an idea and some experimental evidence to this discussion, we focus on one of the most broadly researched topics in both fields, *i.e*., Face Recognition (FR). The latest Neuroscience studies [1, 2, 7, 38, 51] found that

besides the visual cortex, the inferotemporal cortex plays a vital role in the human face-recognizing system. Because of the following three biological characteristics, the inferotemporal cortex characterizes the complicated relationship between attributes and makes attributes contribute to FR. 1) The inferotemporal cortex is functionally compartmentalized by the face stimuli (*i.e.*, *identity*, *attributes*) [1, 7]. 2) The responses of the functional compartments are transformed into the successor neurons for processing complex tasks [38], *e.g.*, FR. 3) The intensities of the functional compartment are variant in the inferotemporal cortex [2], demonstrating that the attributes are not equally essential to the human face-recognizing system. In the AI field, some CNN-based FR works proposed effective loss functions [4, 18, 21, 33, 40, 45] and others designed task-oriented face recognition structures [22, 23, 27–29, 42]. Because of these excellent works, FR in AI achieved big progress. Although a few AI FR works [14, 19, 43] boosted the performance with attribute enhancement, none of them are inspired by the human face-recognizing mechanism nor boosted performance significantly. We experimentally prove that the deep FR models do not capture the biological characteristics of the human face-recognizing system introduced above. We suspect this is the factor limiting their performance improvement.

To address the problem and experimentally support the opinion of Zador, Bengio *et al*. [54], we purposely model the biological characteristics of the human face-recognizing system with classical CNN Ops. We name the proposed **Bio**logically-inspired **Net**work as **BioNet**. The proposed BioNet is constituted by two cascade sub-networks, *i.e.*, **Visual Cortex Network (VCN)** and **Inferotemporal Cortex Network (ICN)**, as Fig.1 shows. The VCN is modeled with a CNN backbone (*e.g.*, CASIA-Net [53], ResNet [12]), which follows the common suggestions in [10, 52]. The proposed ICN is composed of three biologically-inspired modules, *i.e*. the Cortex Functional Compartmentalization (CFC), the Compartment Response Transform (CRT), and the Response Intensity Modulation (RIM). CFC is based on the attention mechanism [8, 15, 44] to functionally compartmentalize the feature maps with face stimuli (*i.e*. *identity* and *attributes*). CRT is implemented by Multilayer Perception (MLP) [36] to transform intra-compartment responses to successor neurons for processing complex FR task. RIM follows the ensemble mechanism [55] to fuse inter-compartment features via adaptive weights to achieve the final FR representation. The proposed modules follow the human face-recognizing mechanism and empower ICN to characterize the complicated relationship between stimuli. All of them are indispensable in BioNet.

With such a Biologically-inspired Network, we achieve better attribute-enhanced deep FR models than ever. Furthermore, we conduct careful analyses of the proposed modules and the impacts of attributes. We also compare the BioNet to the attention, multi-task learning, and ensemble mechanisms, verifying the advantage of our proposals. We think the experiments in this paper can support the conclusions:

1. To the best of our knowledge, after CNN was proposed, our BioNet is the first network inspired by the latest Neuroscience studies. It provides experimental evidence that the latest Neuroscience studies can further boost the CNN-based FR network and continue to drive AI progress.

2. With the Neuroscience mechanism, both identity-related attributes (*e.g.*, *gender*) and identity-unrelated attributes (*e.g.*, *expression*) can benefit the deep FR models. Surprisingly, we find the identity-unrelated ones contribute even more than the identity-related ones. Besides, we also propose an online latent attributes mining method and prove that the latent attributes contribute to the FR task, too.

3. Without bells and whistles, Our BioNet consistently and significantly boosts state-of-the-art of FR models on standard FR benchmark datasets, *e.g.*, IJB-A [24], IJB-B [46], IJB-C [46], and MegaFace [20].

## 2. Related Work

Zador, Bengio, *et al*. [54] claimed that Neuroscience had been a key driver and source of inspiration for improvements in AI, particularly those made AI more proficient in areas that humans and other animals, such as vision, reward-based learning, interacting with the physical world, and language. However, few recent AI studies support their opinion directly. Therefore, some researchers oppose their claim, especially in the CV community. The main questions for those who disagree are: 1) Except for some high-level and abstract senses from Neuroscience, can their specific studies support the CV fields? 2) Since CNN was proposed long ago [9, 26, 52], is there any evidence supporting that the Neuroscience studies released in recent years can continue to drive AI progress? In this paper, we focus on the most broadly studied topic both in CV and Neuroscience communities, *i.e.*, Face Recognition, and attempt to provide evidence to support Zador's opinion experimentally by proposing a Biologically-inspired Face Recognition Network.

**Face-Recognizing System in Neuroscience.** The human face-recognizing system is a topic achieving big progress in Neuroscience recently [1, 2, 7, 38, 51]. Conway *et al*. [2] summarized that besides the visual cortex, the inferotemporal cortex plays a vital role in the human brain. Neuroscience studies also explained how the inferotemporal cortex characterizes the complicated relationship between attributes and makes the attributes contribute to

FR. For example, Freiwald *et al*. [7] and Chang *et al*. [1] discovered that the inferotemporal cortex is functionally compartmentalized by face stimuli, *e.g*., identity and attributes. Moreover, even the compartments stimulated by the identity-unrelated attributes also positively contribute to the face-recognizing system. Conway *et al*. [2] found the intensities of the functional compartment are variant in the inferotemporal cortex. It demonstrates that the attributes are not equally essential to the human face-recognizing system. Pitcher *et al*. [38] concluded that the human brain responses are projected into the deep successor neurons to process complex visual tasks (*e.g*., FR). Although CNN imitated the human primary visual cortex [9,10,52] a long time ago, few CV works are inspired by the latest Neuroscience.

**Face Recognition with CNN.** FR is a typical topic in the CV field. Some CNN-based FR works proposed effective loss functions [4,18,21,33,40,45], *e.g*., AdaFace [21], MagFace [33], and CurricularFace [18]. Others designed task-oriented face recognition structures [22,23,27–29,42], *e.g*., GroupFace [22] and BroadFace [23]. Because of these excellent works, deep FR in the AI field achieved big progress. Although a few AI FR works [14,19,25,30,43] boosted their performance with attribute enhancement, they only employed an inferior backbone and evaluated their proposals in small-scale datasets like LFW [16]. The previous attribute-enhanced FR models are not validated on large-scale challenging evaluation datasets (*e.g*., IJB-B [46], IJB-C [32], and MegaFace [20]). Moreover, NONE of them are inspired by the human face-recognizing system. For example, Kumar *et al*. [25] and Jadhav *et al*. [19] enhanced the traditional no-deep face recognition paradigm (*e.g*., SVM [17] and one-shot learning [48]) with attributes.Hu *et al*. [14] used a fusion matrix to fuse the features from the deep FR network and an individual attribute classification network that are trained separately. Their work did not link to Neuroscience, nor can be trained end-to-end. Taherkhani *et al*. [43] directly fused the feature for attribute classification and the feature after global average pooling to recognize faces. They did not consider the factor that the human inferotemporal cortex is functionally compartmentalized by attributes with variant intensities. Lin *et al*. [30] enhanced Person Re-identification with an attribute re-weighting module. However, they ignore the functional compartmentalization and projection to successor neurons for the complex task in the human face-recognizing system. We experimentally prove the deep FR models do not capture the functional characteristics of the human face-recognizing system introduced before. We suspect this is the factor limiting their performance improvement. There is no previous work that links deep FR with Neuroscience and provides evidence to support that the human face-recognizing system can inspire deep FR community.

# 3. Method: BioNet

In this section, we introduce our proposed BioNet. The BioNet integrates CNN-based models with cutting-edge Neuroscience studies. It consists of two cascade networks, *i.e*., Visual Cortex Network and Inferotemporal Cortex Network, as Fig.1 illustrates.

**Visual Cortex Network (VCN)** is modeled with the CNN backbone directly because LeCun *et al*. [52] and Goodfellow *et al*. [10] theoretically demonstrated that CNN imitates the human visual cortex well.

**Inferotemporal Cortex Network (ICN)** is difficult to be modeled with a vanilla CNN structure, which is proved in Sec.5.3. To this end, we propose three interdepend biologically-inspired modules, *i.e*., CFC, CRT, and RIM, to constitute ICN based on the three essential biological characteristics of the human face-recognizing system. The proposed modules are for:

1) CFC in Sec.3.1 is based on the attention mechanism [8, 15, 44] to compartmentalize the feature maps of VCN by face stimuli.

2) CRT in Sec.3.2 is implemented by MLP [36] to transform intra-compartment features to successor neurons and makes them contribute to FR directly.

3) RIM introduced in Sec.3.3 borrows lessons from the ensemble mechanism [55] to fuse the inter-compartment features based on the adaptive weights to achieve final identification representations.

Experiments demonstrate that the proposed modules capture the biological characteristics of the human face-recognizing system and characterize the complicated relationship of attributes. By integrating cutting-edge Neuroscience studies, our BioNet significantly boosts deep FR performance. It experimentally supports the idea that the latest specific Neuroscience findings can drive CNN-based FR progress.

## 3.1. Cortex Functional Compartmentalization

Neuroscientists, *e.g*., Freiwald *et al*. [7] and Chang *et al*. [1], found that the inferotemporal cortex is functionally compartmentalized by stimuli (*i.e*., identity, attributes) They also discovered that the compartments both stimulated by identity-related and identity-unrelated stimuli contribute to the face-recognizing system. To model this biological characteristic, we propose the Cortex Functional Compartmentalization (CFC) module to compartmentalize the feature maps of VCN by enabling the compartments to classify the stimuli precisely. Specifically, we borrow lessons from the attention mechanism [8, 15, 44] and tune it to a multi-branch supervised-attention mechanism for CFC implementation. Each branch corresponds to a specific stimulus and is supervised by the stimulus classification loss function. The $1^{st}$ stimulus is identity, and the others are

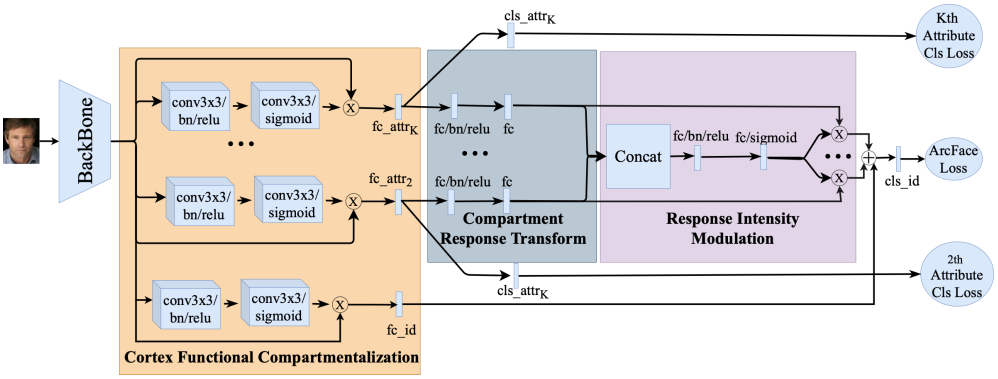


Figure 2. Architecture of the Inferotemporal Cortex Network. It consists of CFC, CRT, and RIM modules.

face attributes. We illustrate CFC in the orange region of Fig.2 and formulate it in the following:

$$\Pi_{cfc_i}(x) = FC_i(x \cdot SA_i(x)), i \in \{1, ..., K\} \quad (1)$$

$x$ is the output of VCN and the input of ICN. $SA_i(x)$ is the $i^{th}$ supervised-attention branch. $FC_i$ is a fully-connected layer. $\Pi_{cfc_i}(x) \in R^{D_i \times 1}$ is the output and the $i^{th}$ functional compartment of CFC. $D_i$ is the feature length. $K$ is the number of stimuli.

Tab.3 demonstrates that CFC successfully functionally compartmentalizes the feature maps in BioNet. Otherwise, the compartments would not classify the attributes precisely without any FR performance degradation.

### 3.2. Compartment Response Transform

Given the functional compartments from CFC, one may directly use them for FR. However, we find in Tab.4 that the functional compartments are helpless or even harmful to FR if used directly. We think it is because the compartments concentrate on attribute classification, which limits their contribution to FR. For improving the FR performance with the latest Neuroscience studies, we model more human biological characteristics to address the problem.

In Neuroscience studies, Pitcher *et al*. [38] found that the compartment responses are transformed into successor neurons for processing complex tasks, *e.g*., FR. We propose the Compartment Response Transform (CRT) to model this biological characteristic. CRT transforms the intra-compartment features to the successor hyper-distribution space and makes them contribute FR directly. In this paper, we implement CRT with a two-layer Multilayer Perception [36], as the gray region in Fig.2 shows. Its formulation is in the following:

$$\Pi_{crt_i}(x) = FC_i^1(FC_i^2(\Pi_{cfc_i}(x)), i \in \{2, ..., K\}) \quad (2)$$

$\Pi_{crt_i}(x) \in \mathbb{R}^{D_1 \times 1}$ is the output of the $i^{th}$ CRT. $D_1$ is the length of identification representation. $\Pi_{cfc_1}(x)$ is optimized for FR by design. Therefore, we do not apply the CRT on it, *i.e*. $\Pi_{crt_1}(x) = \Pi_{cfc_1}(x)$.

The experiments in Tab.4 prove that our implementation is simple yet surprisingly effective. The table also proves CRT is essential to BioNet.

### 3.3. Response Intensity Modulation

Neuroscientists, *e.g*., Conway *et al*. [2], found that the intensities of the functional compartment are variant, demonstrating that the attributes do not equally contribute to FR in the human face-recognizing system. We propose the Response Intensity Modulation (RIM) to model the biological characteristic. RIM modulates the intensities of $\Pi crt(x)$ and fuses inter-compartment responses with the adaptive weights. To estimate the weights across compartments, we are inspired by the ensemble mechanism [55] and concatenate $\Pi_{crt_i}(x), i \in \{2, 3, ..., K\}$ as the input of RIM. With the global concatenating input, RIM is employed via a two-layer Multilayer Perception as Zhang *et al*. [55] did. We illustrate RIM in the purple region of Fig.2 and formulate it as follows:

$$\Pi_{rim}(x) = FC_1(FC_2(Concat(\Pi_{crt_2}(x), , ..., \Pi_{crt_K}(x)))) \quad (3)$$

$\Pi_{rim}(x)_i \in \mathbb{R}$ is the $i^{th}$ output. $\Pi_{rim}(x)_1$ corresponds to the identity stimulus, which is constantly set to $K$.

Fig.3 shows that the impact scores of attributes are variant for FR task in our BioNet, which proves that different attributes contribute differently but positively to FR.

### 3.4. Summary of Inferotemporal Cortex Network

For building the essential human face-recognizing mechanisms with CNN OPs, we input the feature maps of VCN to the CFC module to compartmentalize feature maps via stimuli. The $1^{st}$ stimulus is identity, and the others are attributes. $\Pi_{cfc_i}(x)$ is used to classify the $i^{th}$ attribute that

makes CFC work as our multi-branch supervised-attention mechanism. Simultaneously, $\Pi_{cfc_i}(x)$ is also the input of CRT to get a transformed feature which directly contributes to FR. The final FR representation ,*i.e.*, $feat\_id$, is achieved by ensembling $\Pi_{crt_i}(x)$ with the adaptive outputs of RIM (*i.e.* $\Pi_{rim}(x)_i$), as illustrated in the following Equation:

$$feat\_id = \Pi_{icn}(x) = \frac{1}{K}\sum_{i=1}^{K}\Pi_{rim}(x)_i \times \Pi_{crt_i}(x) \quad (4)$$

Although we borrow lessons from attention, multi-task learning, and ensemble mechanism, Tab.3 proves that the straightforward improvements with the borrowed mechanisms can not capture all three biological characteristics of the human brain as BioNet does. Nor can they boost FR performance as significantly as BioNet. Tab.4 shows that the proposed three biologically-inspired modules are indispensable in ICN.

# 4. Observed and Latent Attributes Annotation

The face attributes annotations are required in our training pipeline. We present two solutions to automatically annotate the observed or the latent attributes:

**Annotating observed attributes offline.** We define objective attributes such as gender, expression, and make-up as the observed attributes. In the paper, we train an attribute classification network to annotate them automatically.

**Annotating latent attributes online.** Kim *et al.* [22] proposed a method to mine the latent groups online. Inspired by their work, we use their group ID to annotate the attributes in our pipeline online. We name the group IDs as latent attributes. For the details on mining latent groups, please refer to their paper [22].

Tab.5 and Tab.6 show that each annotation solution has its advantages. The observed attributes can perform better than the latent ones if the attribute number is equal. It is almost cost-free to increase the number of latent attributes, which makes it easy to improve its performance by increasing the number.

# 5. Experiments

In this section, we prove the effectiveness of BioNet with experiments. Firstly, we compare the FR performance and computational efficiency of our BioNet with the state-of-the-art. Secondly, we show that the vanilla CNN-based FR model and its straightforward improvements do not capture the biological characteristics of the human face-recognizing system. It is quite a possible reason that leads the CNN-based FR difficulty in boosting its performance with attribute knowledge. Thirdly, we study the influence of every proposed module in detail. Fourthly, we carefully analyze the impacts of the latent/observed attributes. Finally, we prove our BioNet is adapted to different CNN backbones, *e.g.* ResNet-101 and CASIA-Net.

## 5.1. Datasets and Implementation Details

**Training Dataset.** We train FR models on the MS-Celeb-1M dataset [11]. We follow the setting in [28] and clean the dataset automatically. The cleaned dataset contains 84,284 identities and 4.8 million images.

**Evaluation Dataset.** We evaluate our BioNet on four of the most challenging datasets in the main paper, *i.e.*, IJB-A [24], IJB-B [46], IJB-C [32], and MegaFace [20] datasets. Additional results on LFW [16], CFP [41], CALFW [57], CPLFW [56], SLLFW [5], and YTF [47] are reported in the supplementary materials because of paper length limitation. **IJB-A** contains 5,397 images and 20,412 video frames from 500 individuals. **IJB-B** extends the IJB-A and contains 1,845 subjects with 21.8K still images and 55K frames from 7,011 videos. **IJB-C** contains 140,740 face images of 3,531 subjects. We report their True Match Rates (TMR) in the paper. Their Rank-K accuracy rates are reported in the supplementary material. **MegaFace** includes the probe and gallery set. The probe set [35] contains 100,000 images of 530 identities, and the gallery set consists of 1,027,060 images from 690,572 subjects. We report the performance on its cleaned dataset [3].

**Attribute Annotations.** With the two presented solutions in Sec.4, we annotated both the observed attributes and latent ones on the MS-Celeb-1M automatically. To annotate the observed attributes offline, we train a CASIA-Net [53] on Celeb-A [31] as the attributes classifier. Because the classifier is only for generating pseudo attribute annotations, we train it with a straightforward strategy as [31] did. The manual annotations noise [49] and unbalanced distributions among attributes [39] in the Celeb-A dataset are not the scopes of this paper. The accuracy of our attributes classifier is reported on the *Baseline:FA* row in Tab.3. Although we conduct an experiment with all forty Celeb-A attributes in Tab.6, we only randomly select four attributes from the forty in other experiments for convenient analysis. Namely, there are five stimuli in most of our experiments, including *identity*, *male* (gender), *mouth-slightly-open* (expression), *smiling* (expression), and *mustache* (make-up). The four selected attributes cover the identity-related attribute (*i.e.*, *male*) and identity-unrelated attributes (*i.e.*, *expression, make-up*). The latent attributes are mined with the proposal in [22].

**Implementation details.** We follow [28] to implement our experiments. The proposed BioNet is trained end-to-end from scratch. The CASIA-Net [53] and ResNet-101 [12] respectively implement the Visual Cortex Network. CASIA-Net randomly samples 96×96 regions from the aligned 100×100 face images for data augmentation. ResNet-101 resizes the face images to 224×224. Both

| | IJB-A | IJB-B | | | IJB-C | | | MegaFace | |
|---|---|---|---|---|---|---|---|---|---|
| | @1e-4 | @1e-4 | @1e-5 | @1e-6 | @1e-4 | @1e-5 | @1e-6 | rank1@1e6 | verify@1e-6 |
| APRN [27] | 94.35 | - | - | - | - | - | - | 98.59 | - |
| AdaFace [21] | - | 95.84 | - | - | 97.09 | - | - | - | - |
| PASS [6] | - | - | - | - | 94.60 | 91.90 | - | - | - |
| ArcFace [4] | - | 94.25 | 89.33 | 38.28 | 96.03 | 93.94 | 89.06 | 98.35 | 98.48 |
| MagFace [33] | - | 94.51 | 90.24 | 42.32 | 95.97 | 94.08 | **90.24** | - | - |
| BroadFace [23] | - | 94.61 | 90.81 | 46.53 | 96.03 | 94.11 | 85.96 | 98.70 | 98.95 |
| CurricularFace [18] | - | 95.83 | 89.02 | 42.26 | 96.20 | 93.85 | 87.46 | 98.71 | 98.64 |
| 3D-BERL [13] | - | 94.98 | 90.60 | 45.77 | 96.20 | 94.30 | 88.45 | 98.63 | 98.64 |
| GroupFace(#group=32) [22] | - | 94.93 | 91.24 | 52.12 | 96.26 | 94.53 | 89.28 | 98.74 | 98.79 |
| Our BioNet: Latent (#attr=4) | 97.27 | **96.13** | 90.66 | 66.09 | 97.19 | 93.89 | 86.38 | 99.03 | **99.80** |
| Our BioNet: Observed (#attr=4) | **97.51** | **96.13** | **92.18** | **68.28** | **97.32** | **94.57** | 88.02 | **99.19** | 99.63 |

Table 1. Comparison with State-of-the-arts. The **bold font** is the best performance.

backbones linearly scale the image intensities to the range $[-1, 1]$. All the feature lengths are 512, *i.e.*, $D_i = 512, i \in \{1, 2, ..., K\}$. We adopt the SGD optimizer with the momentum of 0.9 and weight decay of 0.0005. The learning rate is 0.01 and decays ten times at the $12^{th}$, $14^{th}$, and $15^{th}$ epochs. The total training epoch is set as 16. The FR loss function is the ArcFace loss function [4], and its loss weight is 1. The attributes classification loss function is the Softmax loss function [31], and the sum of attributes classification loss weights is 1, *i.e.*, each of them is $\frac{1}{K-1}$. The networks are trained with the Pytorch [37] on eight NVIDIA V100 GPUs.

## 5.2. Comparison with State-of-the-arts.

We implement the ResNet-101 to the Visual Cortex Network and compare our BioNet with the latest state-of-the-arts, *e.g.*, AdaFace [21], MagFace [33], CurricularFace [18], and GroupFace [22], in Tab.1.The performances of previous attribute-enhanced methods [14, 19, 43] were much inferior to state-of-the-art and were only evaluated in small-scale datasets like LFW [16]. Therefore, we do not compare them in the main paper.

**Recognition Performance Analysis**. Tab.1 shows that our BioNet yields the performance of state-of-the-arts significantly. For example, The performance on IJB-B@1e-6 is improved to 68.28% from 52.12% (GroupFace [22]).Furthermore, it shows BioNet boosted the performance of GroupFace even much more significantly than GroupFace boosted ArcFace [4]. For example, BioNet improved 16.16%/0.45% to GroupFace on IJB-B@1e-6/MegaFace-rank1, while GroupFace only improved 13.84%/0.39% to ArcFace. The comparison to 3D-BERL [13], CurricularFace [18], *etc*. can get the same conclusion.

**Computational Costs Analysis**. The GroupFace [22] is the most related work to ours. Tab.1 shows that BioNet with four attributes has already outperformed the GroupFace [22] with thirty-two groups, which demonstrates that BioNet is more computationally efficient than GroupFace. To prove the claim, we show their Floating Point Operations (FLOPs) [34], the number of parameters (Param), and latency per image in NVIDIA V100 in Tab.2. The table shows that BioNet contains almost the same FLOPs, only 66% latency, and even half parameters to GroupFace.

| | FLOPs | Ratio | Latency | Ratio | Param | Ratio |
|---|---|---|---|---|---|---|
| GroupFace | **7.70 G** | **100%** | 46.75 ms | 100% | 211.7 M | 100% |
| Our BioNet | 8.20 G | 105% | **31.04 ms** | **66%** | **116.2 M** | **55%** |

Table 2. Analysis of computational costs.

## 5.3. Studies about the Vanilla FR Model and Its Straightforward Improvement

For encouraging researchers to nourish the CV field with the latest Neuroscience studies, we purposely model the functional characteristics of the human face-recognizing system with classical CNN Ops. Therefore, we borrow lessons from the attention mechanism, multi-task mechanism, and ensemble mechanism. In Tab.3, we experimentally prove that the straightforward improvements with the borrowed mechanisms can not capture the biological characteristics of the human brain and can not boost FR performance as significantly as BioNet.

**Baseline Implementation**. A CASIA-Net was trained on the MS-Celeb-1M for FR. And then, we freeze all the learnable parameters and train linear classifiers with the features extracted from feat_id/layer4 to classify the selected attributes on Celeb-A. Its performance is shown in the *Baseline:FR*. Besides, another CASIA-Net is trained end-to-end from scratch on Celeb-A as a Face Attribute (FA) classification network and resulted in *Baseline:FA*.

**Straightforward Improvement Implementation.** 1) Multi-task learning mechanism: Both *Multitask I* and *Multitask II* are trained with the multi-task learning mechanism, which import the supervised FA classification tasks. The shared parameters in *Multitask I* are from the start to the last representation layer (*i.e.*, feat_id). The ones shared in *Multitask II* are to the last convolutional layer (*i.e.*, layer4). 2) Attention mechanism: *Self-Attention* employs a self-attention

| | | IJB-A @1e-4 | IJB-B @1e-4 | IJB-C @1e-4 | MegaFace rank1@1e6 | Feat From | Attr4 acc |
|---|---|---|---|---|---|---|---|
| Baseline: CASIA-Net | FA | - | - | - | - | - | 95.29 |
| | FR | 87.39 | 84.16 | 87.40 | 87.01 | feat_id | 70.83 |
| | | | | | | layer4 | 87.60 |
| Straight-forward Improve-ment | Multitask I | 80.96 | 63.02 | 65.47 | 79.67 | feat_id | 93.04 |
| | Multitask II | 87.34 | 84.05 | 87.20 | 86.42 | feat_attrs | 94.92 |
| | Self-Attention | 88.00 | 84.12 | 87.15 | 88.43 | feat_id | 69.01 |
| | | | | | | layer4 | 92.33 |
| | Supervised-Attention | 88.10 | 84.50 | 87.20 | 87.94 | feat_attrs | 95.50 |
| | Avg-Ensemble | 86.93 | 84.29 | 87.17 | 86.92 | feat_attrs | 95.46 |
| | Adaptive-Ensemble | 88.79 | 84.58 | 87.91 | 87.65 | feat_attrs | 95.30 |
| Ours: BioNet | | **89.31** | **85.31** | **88.66** | **90.37** | feat_attrs | **95.54** |

Table 3. Studies about the vanilla FR model and its straightforward improvement. The *feat_id* is the feature for the identification, the *feat_attrs* are the ones for the attribute classifiers, and the *layer4* is the feature obtained by the last convolutional layer.

module without attribute supervision to *layer4*. *Supervised-Attention* employs a multi-branch supervised-attention with attributes supervision. It is similar to our CFC module, but *Supervised-Attention* does not fuse the features to recognize faces. 3) Ensemble mechanism: A single CFC module works as a straightforward *Avg-Ensemble* mechanism because it averages features from multiple tasks to face recognition. Our ICN without CRT works as a straightforward *Adaptive-Ensemble* mechanism because RIM adaptive ensemble features from CFC. The details about all settings are illustrated in supplementary material.

Based on Tab.3, we observe:

1) The vanilla FR model and its straightforward improvements with the borrowed mechanisms can not boost FR performance as significantly as BioNet.

2) The vanilla deep FR model and its straightforward improvements do not capture the biological characteristics of the human inferotemporal cortex. The human inferotemporal cortex is functionally compartmentalized by stimuli, and all compartments contribute to FR performance. However, *Baseline:FR* works poorly in the face attributes classification task (drops almost 10% accuracy rate), and *Multitask I,II* decrease the FR performance. *Self-Attention* and *Supervised-Attention* seem to work well on FR and FA tasks simultaneously, but their improvement of FR is marginal. The performance of self-attention mechanism is comparable to the supervised-attention mechanism. Therefore, we infer that their attribute knowledge does not contribute to FR. *Avg-Ensemble* and *Adaptive-Ensemble* can get the same conclusions as *Self-Attention* and *Supervised-Attention*.

3) Our BioNet boosts the performance of FR with additional attribute knowledge and achieves the best performance on FR and FA tasks. The table also shows that BioNet functionally compartmentalizes the feature maps (otherwise, it could not classify attributes precisely) and makes the attribute compartments contribute to FR (otherwise, it can not be superior to *Baseline* and *Straightforward Improvement* significantly).

The same conclusions also hold on the powerful backbone, *i.e.*, ResNet-101, in the supplementary material.

| Module | CFC | CRT | RIM | IJB-A @1e-4 | IJB-B @1e-4 | IJB-B @1e-5 | IJB-C @1e-4 | IJB-C @1e-5 | MegaFace rank1@1e6 |
|---|---|---|---|---|---|---|---|---|---|
| Baseline: CASIA-Net | | | | 87.39 | 84.16 | 55.30 | 87.40 | 71.75 | 87.01 |
| Module Analysis | ✓ | | | 86.93 | 84.29 | 51.77 | 87.17 | 71.24 | 86.92 |
| | ✓ | ✓ | | 87.97 | 84.93 | 53.04 | 87.66 | 70.68 | 88.42 |
| | ✓ | | ✓ | 88.79 | 84.58 | 58.15 | 87.91 | 74.39 | 87.65 |
| | | ✓ | ✓ | 89.08 | 85.05 | 58.15 | 88.12 | 75.02 | 90.00 |
| BioNet | ✓ | ✓ | ✓ | **89.31** | **85.31** | **58.33** | **88.66** | **75.03** | **90.37** |
| Baseline: ResNet-101 | | | | 96.83 | 95.93 | 85.45 | 97.18 | 91.98 | 98.65 |
| Module Analysis | ✓ | | | 97.10 | 95.77 | 86.64 | 97.04 | 91.83 | 98.83 |
| | ✓ | ✓ | | 97.42 | 96.11 | 86.36 | 97.21 | 91.96 | 99.06 |
| | ✓ | | ✓ | 97.35 | 96.05 | 89.78 | 97.29 | 93.14 | 98.89 |
| | | ✓ | ✓ | 97.46 | 96.12 | 90.14 | 97.32 | 93.52 | 98.89 |
| BioNet | ✓ | ✓ | ✓ | **97.51** | **96.13** | **92.18** | **97.32** | **94.57** | **99.19** |

Table 4. Ablation studies about the proposed modules.

## 5.4. Ablation Studies about the Proposed Modules

We do ablation studies with four observed attributes to analyze the proposed modules. As Tab.4 presents:

1) BioNet boosts the Baseline on all evaluation datasets significantly. For example, the performance of ResNet-101 on MegaFace is improved to $99.19\%$ from $98.65\%$.

2) The first row in *Module Analysis* shows that the individual CFC seems helpless or even harmful to FR. However, CFC conducts functional compartmentalization and is the precondition of the other two modules. Removing it will limit the performance improvement of BioNet, as the last row in *Module Analysis* shows.

3) An improvement degradation is observed when we remove RIM on the second row of *Module Analysis*. Balancing the impacts of attribute compartments for FR is necessary.

4) Removing CRT leads to improvement degradation on FR, as the third row of *Module Analysis* shows. It is important to transform the attribute features before applying them to face recognition.

Overall, all three proposed modules are indispensable.

## 5.5. Ablation Studies about the Attributes

In this sub-section, we first analyze the influence of observed attributes and latent attributes. Then, the impacts of observed attribute categories are explored. Finally, we analyze the influence of attribute number.

**Influence of observed attributes and latent attributes** is shown in Tab.5. Both the observed attribute number and the latent attribute number are equal to four. To analyze the necessity of attribute annotations, we also implement a BioNet$^-$ without the attribute stimuli, *i.e.*, CFC in BioNet$^-$ is not supervised by attributes annotations.

| | Attr Type | IJB-A | IJB-B | | IJB-C | | MegaFace |
|---|---|---|---|---|---|---|---|
| | | @1e-4 | @1e-4 | @1e-5 | @1e-4 | @1e-5 | rank1@1e6 |
| Baseline: CASIA-Net | | 87.39 | 84.16 | 55.30 | 87.40 | 71.75 | 87.01 |
| BioNet$^-$ | None | 88.77 | 85.08 | 45.15 | 88.12 | 63.63 | 89.69 |
| BioNet | Latent | 88.30 | **86.00** | 55.32 | 88.60 | 73.47 | 90.32 |
| | Observed | **89.31** | 85.31 | **58.33** | **88.66** | **75.03** | **90.37** |
| Baseline: ResNet-101 | | 96.83 | 95.93 | 85.45 | 97.18 | 91.98 | 98.65 |
| BioNet$^-$ | None | 96.82 | 95.96 | 88.05 | 97.12 | 92.56 | 99.02 |
| BioNet | Latent | 97.27 | **96.13** | 90.66 | 97.19 | 93.89 | 99.03 |
| | Observed | **97.51** | **96.13** | **92.18** | **97.32** | **94.57** | **99.19** |

Table 5. Analysis of the observed and latent attributes.

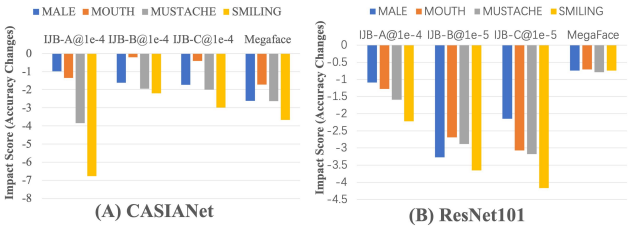


Figure 3. Impacts of the observed attribute categories.

Tab.5 shows that both observed and latent attributes improve performance. Besides, the improvement of observed attributes is more significant than the latent ones, which demonstrates that the observed attributes are superior to the latent ones if their numbers are equal. Furthermore, the improvement of BioNet$^-$ is NOT as significant as BioNet, which proves the attributes stimuli are essential.

**Impacts of the observed attribute categories** are illustrated in Fig.3 via impact score [50]. In our experiments, the impact score of $i^{th}$ attribute is the performance changes when $\Pi_{crt_i}(x)$ is removed in the inference phase, *i.e.*, $impact\_score_i = Accuracy\_without\_attr_i - Accuracy$. A negative score means removing the attribute would lead negative impact. The absolute value is the intensity of impact. For the details of impact score, please refer to [50].

In Fig.3, we find that removing any attributes will decrease FR performance. It proves that all the attribute compartments contribute to FR. Furthermore, we surprisingly find that the most important attribute for FR is not the identity-related attribute (*i.e.*, *male*) but the identity-unrelated one (*i.e.*, *smiling*). For example, employing the ResNet101 to VCN, the performance on the IJB-C@1e-5 dataset decreases by 4.17% when the *smiling* is removed, but the impact score of *male* is $-2.15\%$. We think it is because parts of knowledge from the compartments that correspond to the identity-related attributes have already been contained in the identity compartments. The negative impact is limited when removing duplicate knowledge.

**Influence of the attribute number (K)** is explored in Tab.6. We study the influence of attributes number via latent attributes because they can vanish the interference of the attribute category. Besides, we also employ the selected four and all forty observed attributes as comparisons.

| | #attr | IJB-A | IJB-B | | IJB-C | | MegaFace |
|---|---|---|---|---|---|---|---|
| | | @1e-4 | @1e-4 | @1e-5 | @1e-4 | @1e-5 | rank1@1e6 |
| Baseline: CASIA-Net | 0 | 87.39 | 84.16 | 55.30 | 87.40 | 71.75 | 87.01 |
| BioNet: Observed | 4 | 89.31 | 85.31 | 58.33 | 88.66 | 75.03 | 90.37 |
| | 40 | **90.52** | **87.75** | **69.46** | **89.95** | **80.29** | **94.92** |
| BioNet: Latent | 4 | 88.30 | 86.00 | 55.32 | 88.60 | 73.47 | 90.32 |
| | 10 | 88.72 | 86.11 | 63.63 | 88.84 | 77.53 | 91.39 |
| | 16 | 88.36 | 86.58 | 65.27 | 89.02 | 77.53 | 91.91 |
| | 22 | 89.06 | 86.45 | 65.68 | 89.90 | 78.37 | 92.23 |
| | 28 | 89.01 | 87.20 | 66.72 | 89.31 | 77.91 | 92.32 |
| | 34 | 90.16 | 87.00 | 65.28 | 89.51 | 78.40 | 92.54 |
| | 40 | 90.34 | 87.74 | 65.79 | 89.25 | 78.39 | 92.80 |

Table 6. Influence of attribute number (K).

Tab.6 shows that the performance is improved in every increased step of attribute number.The experiments whose VCN is ResNet-101 in the supplementary material also support the conclusion.

### 5.6. Ablation Studies about CNN Backbones

Tab.5 shows our BioNet is adapted to different backbones, *e.g.*, CASIA-Net and ResNet-101. Besides, we observe that ResNet-101 is improved more significantly by our BioNet than CASIA-Net. For example, the performance of ResNet-101 on IJB-C@1e-5 is improved to 94.57% from 91.98%, while the CAISA-Net is from 71.75% to 75.03%. We infer the bottleneck that limits performance improvement is the inferiority of CASIA-Net.

## 6. Conclusion and Future Work

In this paper, we experimentally support the opinion that the latest Neuroscience can inspire CV progress. Therefore, we purposely model the functional characteristics of the human face-recognizing system with classical CNN Ops. The proposed BioNet consistently and significantly boosts state-of-the-art face recognition methods. Some observations are also discovered. For example, the observed attributes contribute more to FR than the latent attributes, and identity-unrelated attributes (*e.g.*, *smiling*) contribute more than the identity-related ones (*e.g.*, *male*). We believe more interesting discoveries can be made by integrating Neuroscience with CV development.

While these initial results on FR are encouraging, many challenges remain. One is applying Neuroscience studies to other CV tasks, *e.g.*, image classification, object detection, *etc*. Another challenge is integrating Neuroscience into other deep paradigms, *e.g.*, Transformer [44]. In fact, BioNet and Transformer share some similar ideas. For example, they all borrow lessons from the attention mechanism and project responses into successor neurons via CRT or FFN [44]. It could be a potential and exciting topic to improve Transformer with Neuroscience studies.

**Potential negative impact.** The abuse on military applications and the potential privacy issues are two major negative impacts of our BioNet. Therefore, careful evaluations should be conducted before applying it in real applications.

## References


[1] Le Chang and Doris Y Tsao. The code for facial identity in the primate brain. *Cell*, 169(6):1013–1028, 2017. 1, 2, 3

[2] Bevil R Conway. The organization and operation of inferior temporal cortex. *Annual review of vision science*, 4:381–402, 2018. 1, 2, 3, 4

[3] JianKang Deng and Jia Guo. https://github.com/deepinsight/insightface, 2019. 5

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019. 2, 3, 6

[5] Weihong Deng, Jiani Hu, Nanhai Zhang, Binghui Chen, and Jun Guo. Fine-grained face verification: Fglfw database, baselines, and human-dcmn partnership. *Pattern Recognition*, 66:63–73, 2017. 5

[6] Prithviraj Dhar, Joshua Gleason, Aniket Roy, Carlos D Castillo, and Rama Chellappa. Pass: Protected attribute suppression system for mitigating bias in face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15087–15096, 2021. 6

[7] Winrich A Freiwald and Doris Y Tsao. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005):845–851, 2010. 1, 2, 3

[8] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019. 2, 3

[9] Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982. 1, 2, 3

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. 1, 2, 3

[11] 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. 5

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 630–645. Springer, 2016. 2, 5

[13] Mingjie He, Jie Zhang, Shiguang Shan, and Xilin Chen. Enhancing face recognition with self-supervised 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4062–4071, 2022. 6

[14] Guosheng Hu, Yang Hua, Yang Yuan, Zhihong Zhang, Zheng Lu, Sankha S Mukherjee, Timothy M Hospedales, Neil M Robertson, and Yongxin Yang. Attribute-enhanced face recognition with neural tensor fusion networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3744–3753, 2017. 2, 3, 6

[15] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 2, 3

[16] 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. 3, 5, 6

[17] Jeffrey Huang, Xuhui Shao, and Harry Wechsler. Face pose discrimination using support vector machines (svm). In *Proceedings. fourteenth international conference on pattern recognition (Cat. No. 98EX170)*, volume 1, pages 154–156. IEEE, 1998. 3

[18] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2020. 2, 3, 6

[19] Aishwarya Jadhav, Vinay P Namboodiri, and KS Venkatesh. Deep attributes for one-shot face recognition. In *European Conference on Computer Vision*, pages 516–523. Springer, 2016. 2, 3, 6

[20] 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 (CVPR)*, pages 4873–4882, 2016. 2, 3, 5

[21] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18750–18759, 2022. 2, 3, 6

[22] Yonghyun Kim, Wonpyo Park, Myung-Cheol Roh, and Jongju Shin. Groupface: Learning latent groups and constructing group-based representations for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5621–5630, 2020. 2, 3, 5, 6

[23] Yonghyun Kim, Wonpyo Park, and Jongju Shin. Broadface: Looking at tens of thousands of people at once for face recognition. In *European Conference on Computer Vision*, pages 536–552. Springer, 2020. 2, 3, 6

[24] Brendan F Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1931–1939, 2015. 2, 5

[25] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. Attribute and simile classifiers for face verification. In *2009 IEEE 12th international conference on computer vision*, pages 365–372. IEEE, 2009. 3

[26] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995. 1, 2

[27] Pengyu Li, Biao Wang, and Lei Zhang. Adversarial pose regression network for pose-invariant face recognitions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1940–1948, 2021. 2, 3, 6

[28] Pengyu Li, Biao Wang, and Lei Zhang. Virtual fully-connected layer: Training a large-scale face recognition dataset with limited computational resources. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13315–13324, 2021. 2, 3, 5

[29] Wenyu Li, Tianchu Guo, Pengyu Li, Binghui Chen, Biao Wang, Wangmeng Zuo, and Lei Zhang. Virface: Enhancing face recognition via unlabeled shallow data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14729–14738, 2021. 2, 3

[30] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. Improving person re-identification by attribute and identity learning. *Pattern Recognition*, 95:151–161, 2019. 3

[31] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015. 5, 6

[32] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165. IEEE, 2018. 3, 5

[33] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14225–14234, 2021. 2, 3, 6

[34] Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila, and Jan Kautz. Pruning convolutional neural networks for resource efficient inference. *arXiv preprint arXiv:1611.06440*, 2016. 6

[35] Hong-Wei Ng and Stefan Winkler. A data-driven approach to cleaning large face datasets. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 343–347. IEEE, 2014. 5

[36] Leonardo Noriega. Multilayer perceptron tutorial. *School of Computing. Staffordshire University*, 2005. 2, 3, 4

[37] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[38] David Pitcher and Leslie G Ungerleider. Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, 2020. 1, 2, 3, 4

[39] Ethan M Rudd, Manuel Günther, and Terrance E Boult. Moon: A mixed objective optimization network for the recognition of facial attributes. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*, pages 19–35. Springer, 2016. 5

[40] 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. 2, 3

[41] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–9. IEEE, 2016. 5

[42] Yichun Shi, Xiang Yu, Kihyuk Sohn, Manmohan Chandraker, and Anil K Jain. Towards universal representation learning for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6817–6826, 2020. 2, 3

[43] Fariborz Taherkhani, Nasser M Nasrabadi, and Jeremy Dawson. A deep face identification network enhanced by facial attributes prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 553–560, 2018. 2, 3, 6

[44] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017. 2, 3, 8

[45] 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. 2, 3

[46] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 90–98, 2017. 2, 3, 5

[47] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534. IEEE, 2011. 5

[48] Lior Wolf, Tal Hassner, and Yaniv Taigman. The one-shot similarity kernel. In *2009 IEEE 12th International Conference on Computer Vision*, pages 897–902. IEEE, 2009. 3

[49] Haiyu Wu, Grace Bezold, Manuel Günther, Terrance Boult, Michael C King, and Kevin W Bowyer. Consistency and accuracy of celeba attribute values. *arXiv preprint arXiv:2210.07356*, 2022. 5

[50] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1249–1258, 2016. 8

[51] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016. 1, 2

[52] Le Cun Yan, B Yoshua, and H Geoffrey. Deep learning. *nature*, 521(7553):436–444, 2015. 1, 2, 3

[53] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014. 2, 5

[54] Anthony Zador, Blake Richards, Bence Ölveczky, Sean Escola, Yoshua Bengio, Kwabena Boahen, Matthew Botvinick, Dmitri Chklovskii, Anne Churchland, Claudia Clopath, et al.

Toward next-generation artificial intelligence: Catalyzing the neuroai revolution. *arXiv preprint arXiv:2210.08340*, 2022. 1, 2

[55] Weihe Zhang, Yali Wang, and Yu Qiao. Metacleaner: Learning to hallucinate clean representations for noisy-labeled visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7373–7382, 2019. 2, 3, 4

[56] Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5, 2018. 5

[57] 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. 5