DAIL: Dataset-Aware and Invariant Learning for Face Recognition

Gaoang Wang[1,2], Lin Chen[1], Tianqiang Liu[1], Mingwei He[1], and Jiebo Luo[3]

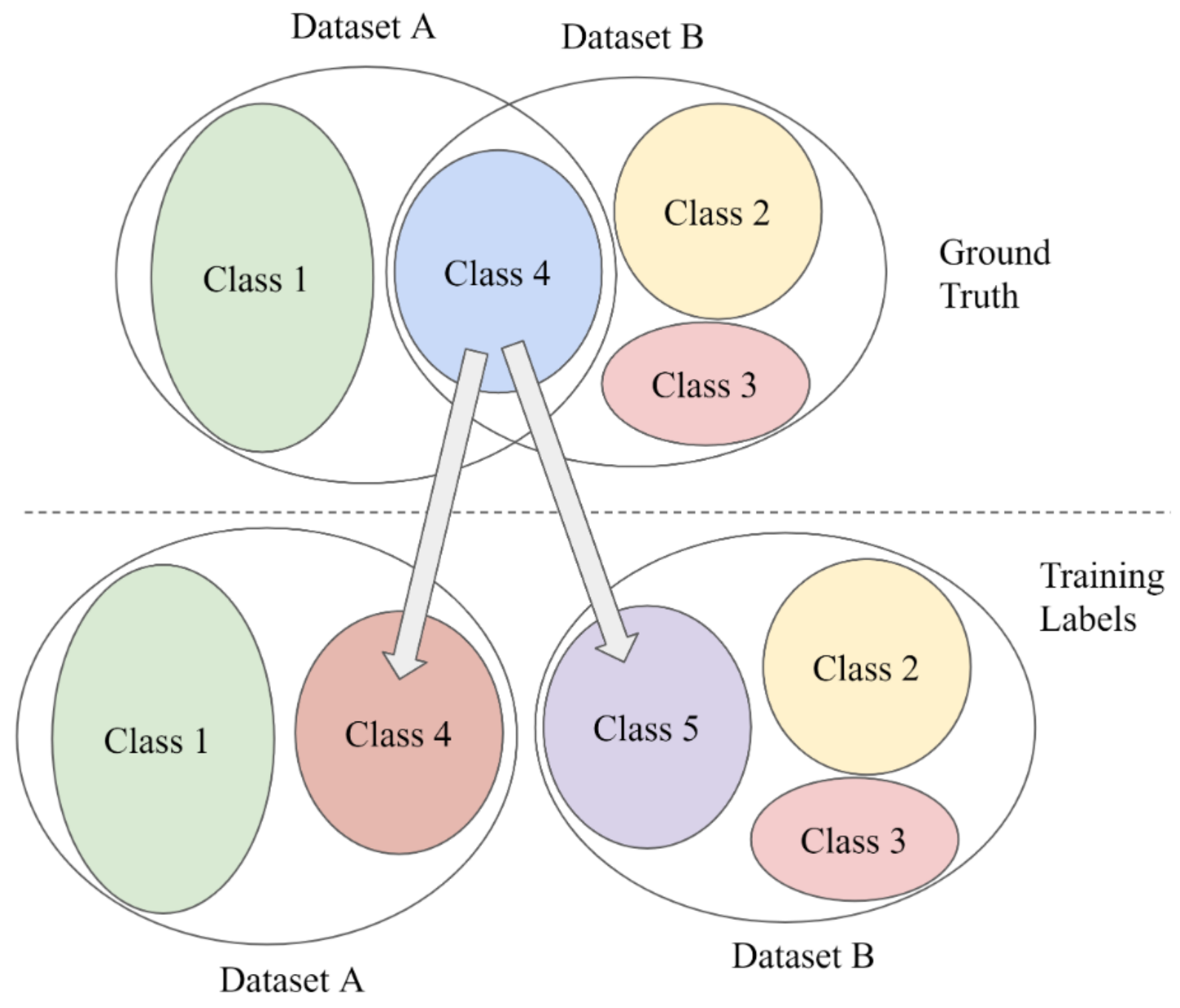
[1]Wyze Labs, Kirkland, WA 98033, USA
[2]Zhejiang University / University of Illinois at Urbana-Champaign Institute, Haining, Zhejiang 314400, China
[3]University of Rochester, Rochester, NY 14627, USA
Email: gaoangwang@intl.zju.edu.cn,{lchen, tliu, mhe}@wyze.com, jluo@cs.rochester.edu

Abstract

To achieve good performance in face recognition, a large scale training dataset is usually required. A simple yet effective way to improve the recognition performance is to use a dataset as large as possible by combining multiple datasets in the training. However, it is problematic and troublesome to naively combine different datasets due to two major issues. First, the same person can possibly appear in different datasets, leading to an identity overlapping issue between different datasets. Naively treating the same person as different classes in different datasets during training will affect back-propagation and generate non-representative embeddings. On the other hand, manually cleaning labels may take formidable human efforts, especially when there are millions of images and thousands of identities. Second, different datasets are collected in different situations and thus will lead to different domain distributions. Naively combining datasets will make it difficult to learn domain invariant embeddings across different datasets. In this paper, we propose DAIL: Dataset-Aware and Invariant Learning to resolve the above-mentioned issues. To solve the first issue of identity overlapping, we propose a dataset-aware loss for multi-dataset training by reducing the penalty when the same person appears in multiple datasets. This can be readily achieved with a modified softmax loss with a dataset-aware term. To solve the second issue, domain adaptation with gradient reversal layers is employed for dataset invariant learning. The proposed approach not only achieves the state-of-the-art results on several commonly used face recognition validation sets, including LFW, CFP-FP, and AgeDB-30, but also shows great benefit for practical use.

Issues for Multi-Dataset Training

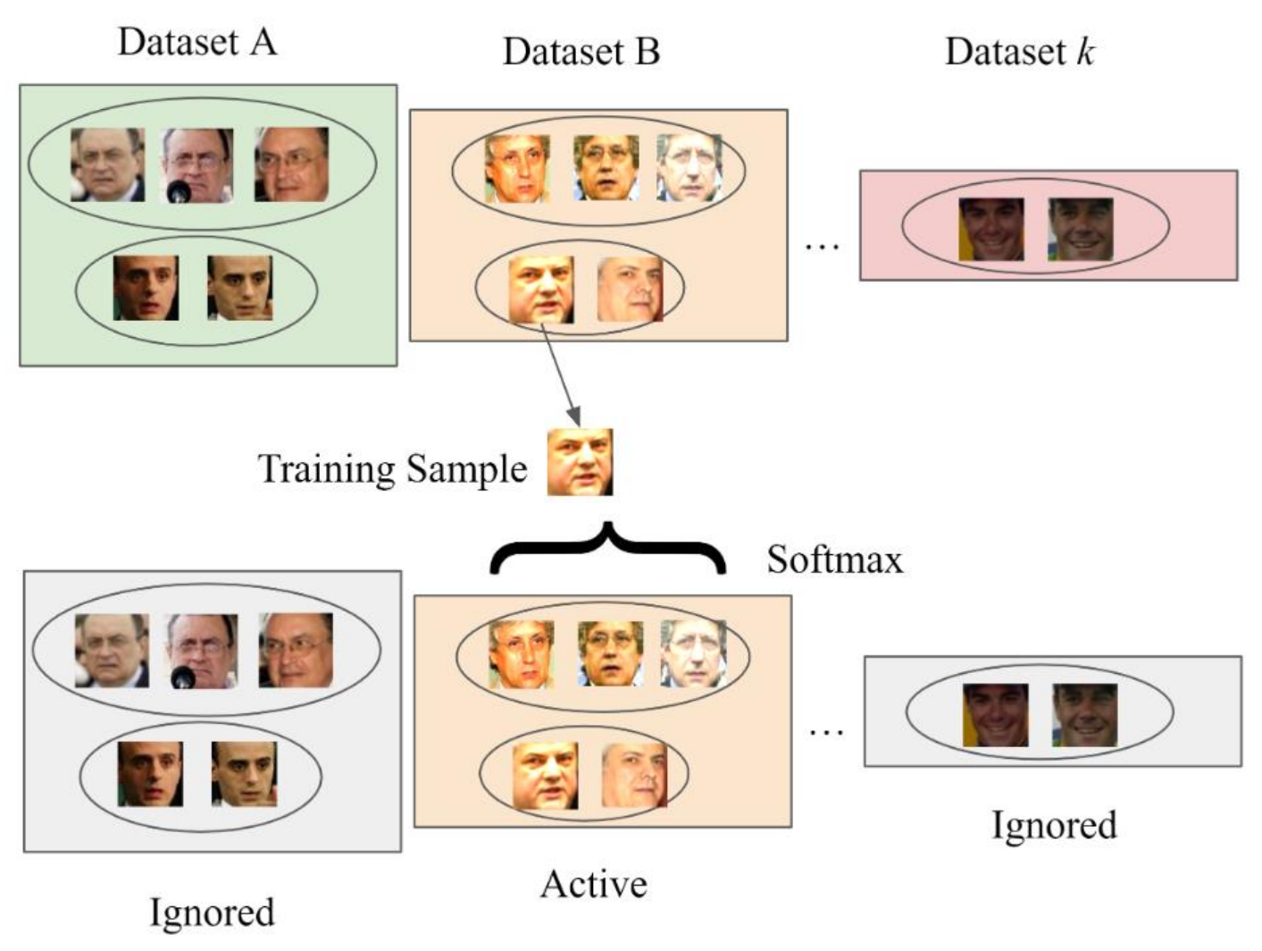
- Labels overlap across datasets
  - Harmful for training
  - Cleaning is expensive



Dataset-Aware Loss

- Definition

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\mathbf{W}_{y_i}^T\mathbf{x}_i+b_{y_i}}}{e^{\mathbf{W}_{y_i}^T\mathbf{x}_i+b_{y_i}}+\sum_{j=1,j\neq y_i}^{C}\mathbf{1}_{k_j=k_{y_i}}e^{\mathbf{W}_j^T\mathbf{x}_i+b_j}}$$

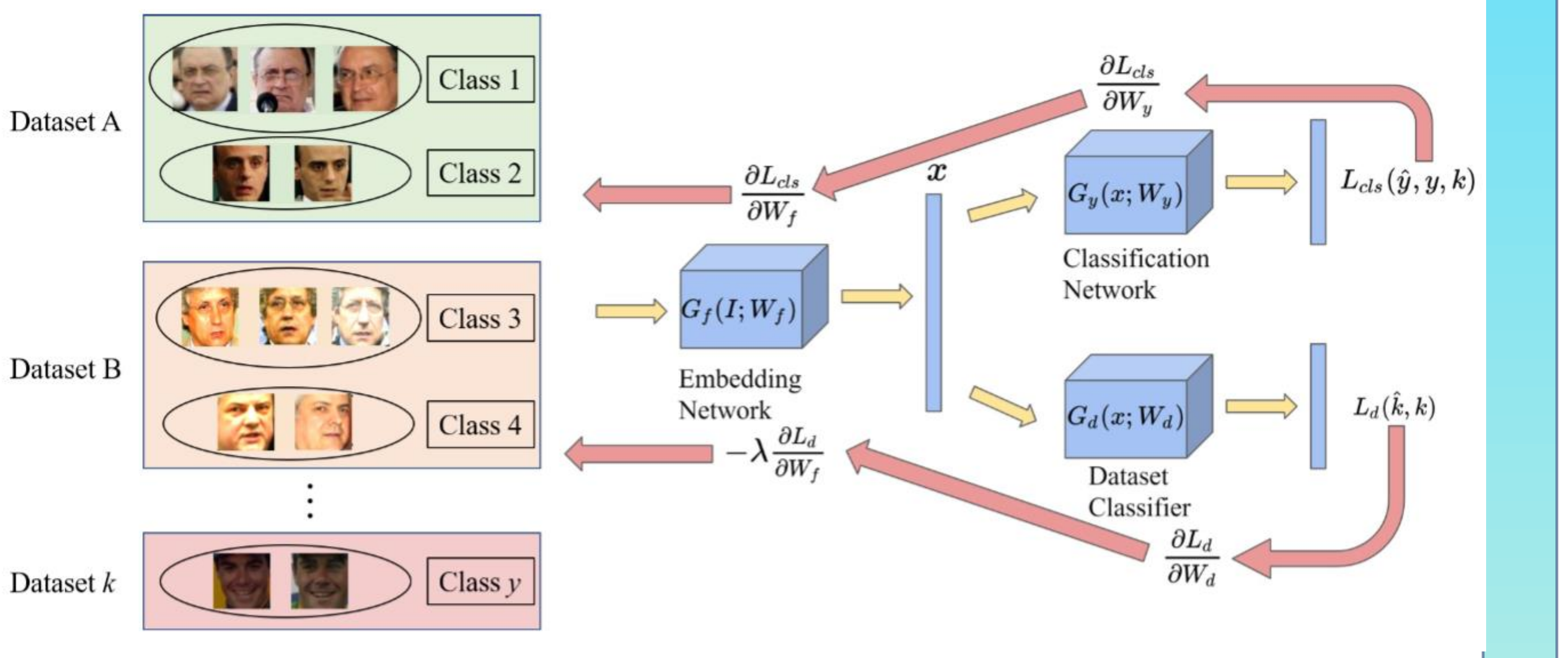


- Benefits
  - Solve label overlapping issues
  - Easy to be combined with other losses, like ArcFace [1].

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)}+\sum_{j=1,j\neq y_i}^{C}\mathbf{1}_{k_j=k_{y_i}}e^{s\cos\theta_j}}$$

Dataset-Invariant Learning by Domain Adaptation

- Learn with gradient reversal layers (GRL) [2]
  - Learn feature embeddings from different domains (datasets)



- Total loss and training

$$\begin{aligned}
L_{cls} &= \sum_i J_{cls}(G_y(G_f(I_i;\mathbf{W}_f);\mathbf{W}_y),y_i,k_i)\\
&= \sum_i J_{cls}(G_y(\mathbf{x}_i;\mathbf{W}_y),y_i,k_i),\\
L_d &= \sum_i J_d(G_d(G_f(I_i;\mathbf{W}_f);\mathbf{W}_d),k_i)\\
&= \sum_i J_d(G_d(\mathbf{x}_i;\mathbf{W}_d),k_i).
\end{aligned}$$

$$\begin{aligned}
(\hat{\mathbf{W}}_f,\hat{\mathbf{W}}_y) &= \underset{\mathbf{W}_f,\mathbf{W}_y}{\operatorname{argmin}}\ \Big\{L_{cls}(\mathbf{W}_f,\mathbf{W}_y,y,k)\\
&\qquad - \lambda L_d(\mathbf{W}_f,\hat{\mathbf{W}}_d,k)\Big\},\\
\hat{\mathbf{W}}_d &= \underset{\mathbf{W}_d}{\operatorname{argmin}}\ L_d(\hat{\mathbf{W}}_f,\mathbf{W}_d,k).
\end{aligned}$$

Experiments

- Datasets
  - Training
    - 14-Celebrity, Asian-Celeb, CASIA, CelebA, DeepGlint, MS1M, PinsFace, 200-Celeb, VGGFace2 and UMDFace
  - Validation
    - LFW, CFP-FP and AgeDB-30

Dataset	#ID	#Image
14-Celebrity [50]	14	117
Asian-Celeb [51]	94.0K	2.8M
CASIA [25]	10.5K	0.5M
CelebA [52]	10.2K	0.2M
DeepGlint [51]	180.9K	6.8M
MS1M [23]	85.7K	5.8M
PinsFace [53]	105	14.1K
200-Celeb	268	24.9K
VGGFace2 [22]	8.6K	3.1M
UMDFace [54]	8.3K	0.4M
LFW [55]	5.7K	13,233
CFP-FP [56]	500	7,000
AgeDB-30 [57]	568	16,488

- Verification accuracy with different losses

Loss	Dataset	LFW	CFP-FP	AgeDB-30
SphereFace	CASIA	99.1	94.4	91.7
CosFace	CASIA	99.5	95.4	94.6
CM (0.9, 0.4, 0.15)	CASIA	99.5	95.2	94.9
ArcFace	CASIA	99.5	95.6	95.2
ArcFace	MS1M	99.8	92.7	97.8
Proposed	Comb	99.8	98.7	98.2

[a]All models are using ResNet50 for embedding.

- Verification accuracy on LFW with SOTA methods

Method	#Image	LFW
DeepID [21]	0.2M	99.5
Deep Face [58]	4.4M	97.4
VGG Face [13]	2.6M	99.0
FaceNet [10]	200M	99.6
Baidu [59]	1.3M	99.1
Center Loss [6]	0.7M	99.3
Range Loss [11]	5M	99.5
Marginal Loss [12]	3.8M	99.5
Proposed (ResNet50)	19.6M	99.8

- Ablation study for different components

Method	Dataset	LFW	CFP-FP	AgeDB-30
ArcFace	MS1M	99.5	88.9	95.9
ArcFace	VGGFace2	99.5	94.2	93.6
Naive Comb	Comb	99.1	95.0	94.8
DA	Comb	99.5	95.4	95.7
DA+GRL	Comb	99.7	96.0	96.0
DA+GRL+CD	Comb	99.6	96.3	96.2

DA: dataset-aware loss
GRL: gradient reversal layer
CD: crossing dropout

$$\mathbf{1}_{k_i=k_j,z<p} = \begin{cases} 1, & \text{if } k_i = k_j\text{, or } k_i \neq k_j \text{ and } z < p,\\ 0, & \text{otherwise,}\end{cases}$$

References

- [1] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4690-4699).
- [2] Ganin, Y., & Lempitsky, V. (2015, June). Unsupervised domain adaptation by backpropagation. In International conference on machine learning (pp. 1180-1189). PMLR.