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

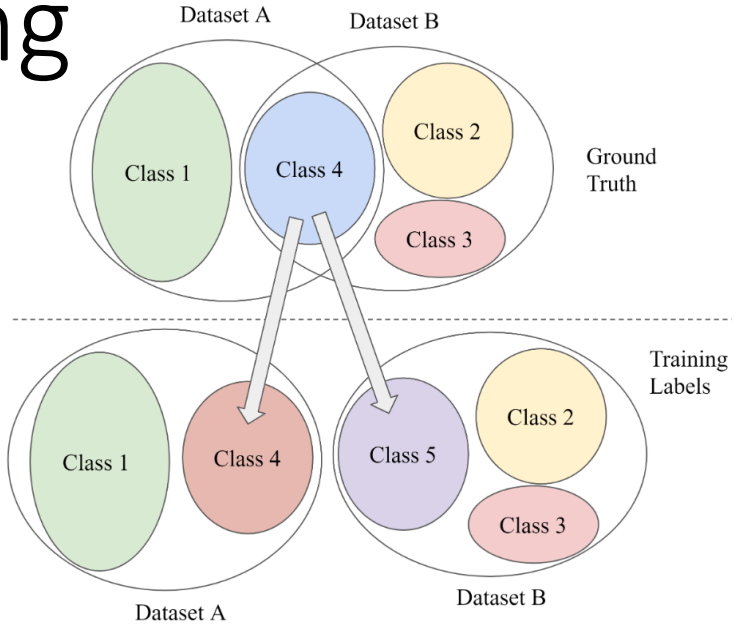
Email: gaoangwang@intl.zju.edu.cn,{lchen, tliu, mhe}@wyze.com, jluo@cs.rochester.edu

Introduction

- Background
  - Good model for face recognition requires large amount of labeled data
  - Data labeling is time-consuming and expensive
  - There are lots of datasets created for face recognition: Asian-Celeb, CASIA, DeepGlint, PinsFace, 200-Celeb, VGGFace2, UMDFace, etc.
- How to become even better?
  - Combine multiple datasets

Issues on multi-dataset training

- Label overlapping
  - Same person appears in multiple datasets
  - Harmful for training
  - Data cleaning is expensive if not impossible
- Dataset bias
  - Different datasets are collected in different time or in a different manner with distribution mismatch



DAIL: **D**ataset **A**ware and **I**nvariant **L**earning

- Dataset-aware loss
  - Solve label overlapping issues without data cleaning
- Gradient reversal layers (GRL) [1] for domain adaptation
  - Learn dataset invariant feature embeddings

[1] Ganin, Y., & Lempitsky, V. (2015, June). Unsupervised domain adaptation by backpropagation. In *International conference on machine learning* (pp. 1180-1189). PMLR.

Dataset-aware loss

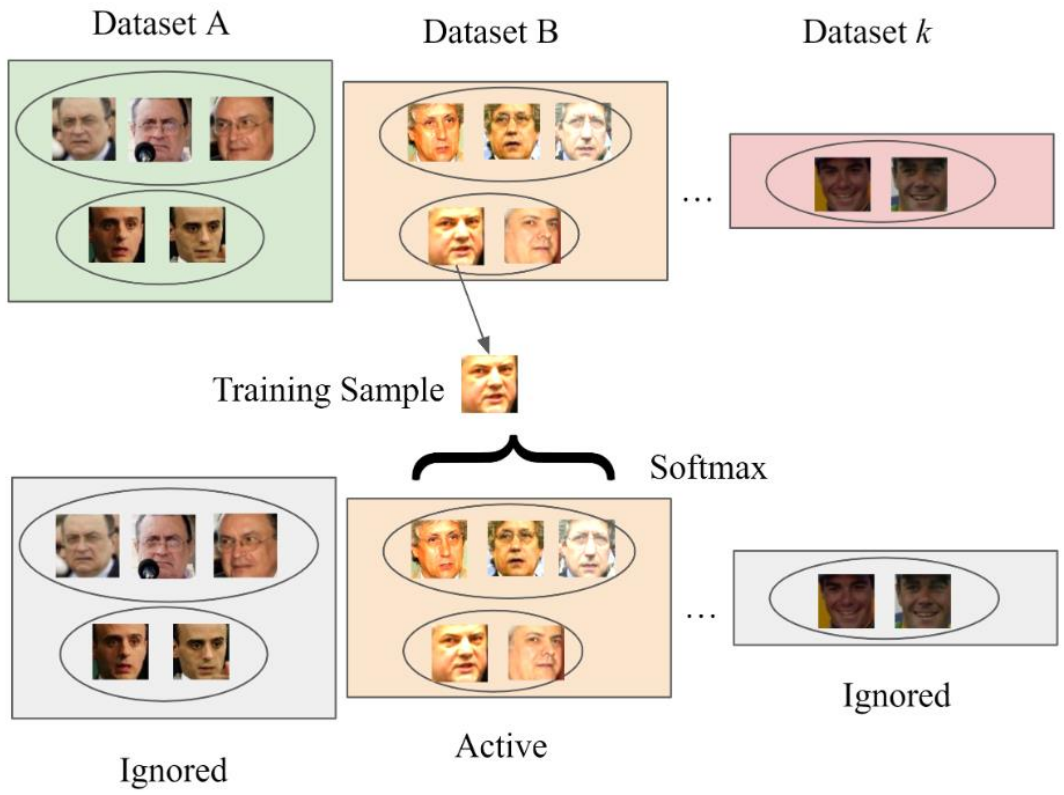
- Definition

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

Consider the same dataset

- Benefits
  - Solve label overlapping issues
  - Easy to be combined with other softmax losses, like ArcFace [2]

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



[2] 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).

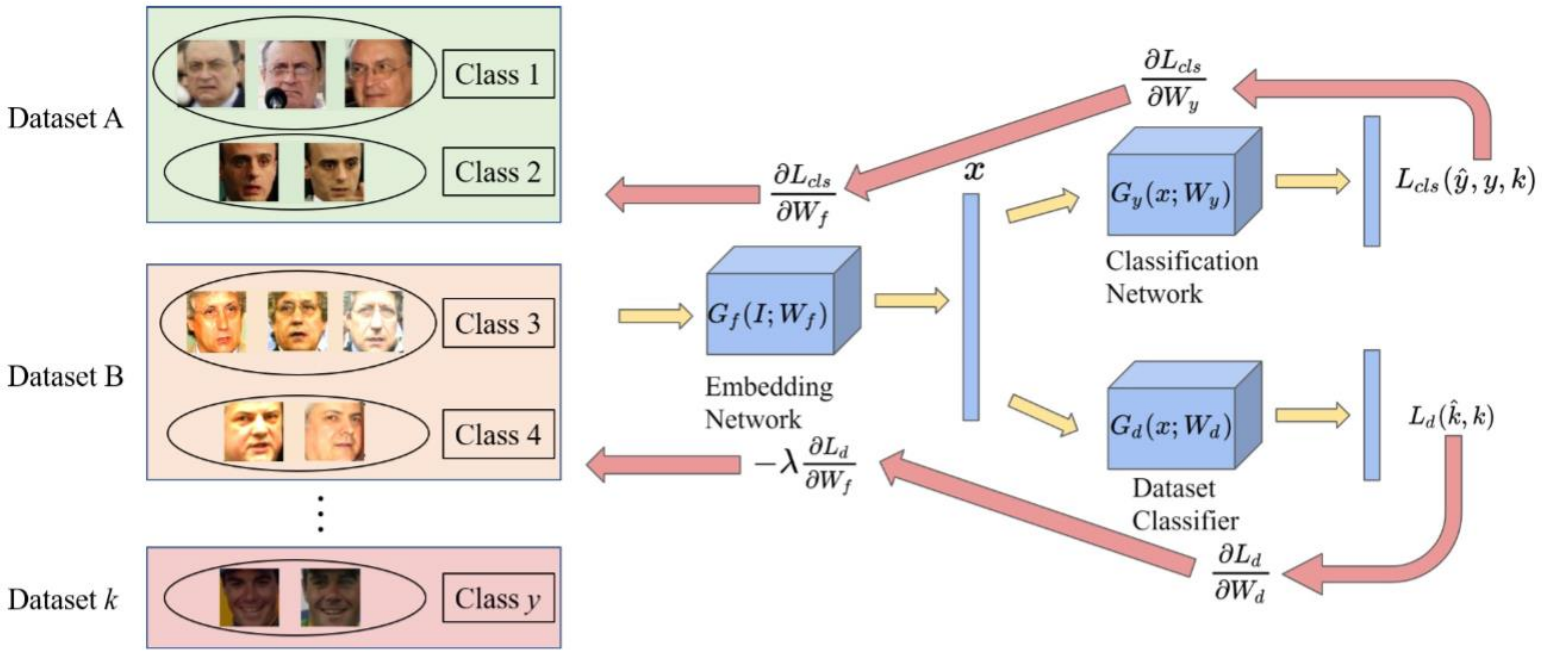
Dataset-invariant learning by domain adaptation

- Train with gradient reversal layers (GRL)
  - Learn feature embeddings from different domains (datasets)

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

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

$$
\hat{\mathbf{W}}_d = \underset{\mathbf{W}_d}{\operatorname{argmin}} \; L_d(\hat{\mathbf{W}}_f, \mathbf{W}_d, k).
$$



Experiments: datasets

- Training datasets
  - 14-Celebrity, Asian-Celeb, CASIA, CelebA, DeepGlint, MS1M, PinsFace, 200-Celeb, VGGFace2 and UMDFace.
- Validation datasets
  - 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

Experiments: 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.

Experiments: 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

Experiments: ablation study for different components

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

DA: dataset-aware loss
GRL: gradient reversal layer
CD: crossing dropout with p=0.0001

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

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

All models are using MobileNet for embedding

Conclusions & take-home messages

- Good face recognition model relies on large amount of high-quality labeled data
- Natively combining multiple datasets for training face recognition model can be harmful because of label overlapping and dataset bias
- Our proposed DAIL solves the label overlapping issue with a dataset-aware loss and dataset bias issue with domain adaptation using GRL
- DAIL is simple but effective and can be used for training production model where multiple datasets are collected in different time and setups

Thanks ☺