LARNet: Lie Algebra Residual Network for Face Recognition

Xiaolong Yang, Xiaohong Jia, Dihong Gong, Dong-Ming Yan, Zhifeng Li, Wei Liu

AMSS,CAS

UCAS

Tencent Data Platform

CASIA

Introduction

- A major challenge in practical face recognition applications lies in significant variations between profile and frontal faces.

NOT EASY for face recognition

with large pose or profile faces !

Introduction

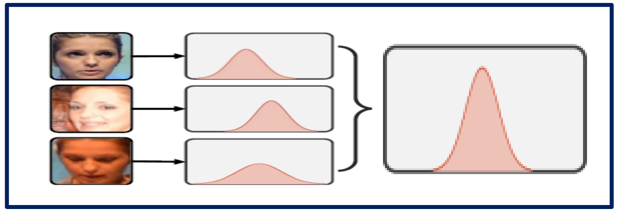
- Traditional techniques:

Masi et al. (2016)

- Enrich sources by the synthesis of profile faces
- Unnecessary computational burdens

Rotate-and-Render (2020)

- Frontalize faces under all pose ranges
- Occlusions and non-rigid expressions

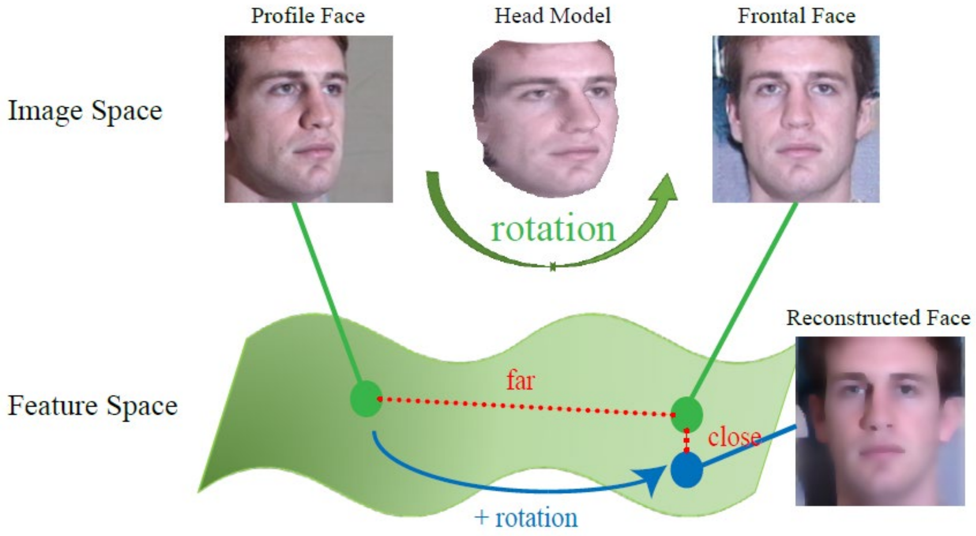


Probabilistic Face Embeddings (2019)

- Represent face images as a Gaussian distribution
- Overly strong hypothesis prior

Our method

- We explore how face rotation in the 3D space affects the deep feature generation process.



The difference between frontal and profile face images

→ Head rotation in 3D space

Face recognition for images

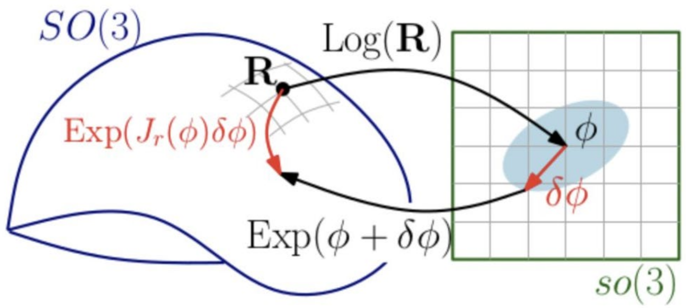
→ Clustering in the deep feature space

Our method

- Rotation matrices are not easy to be embedded in CNNs.
  - Rotation group is not completely closed.

$\forall R_1, R_2 \in SO(3),\ R_1 \cdot R_2 \in SO(3)$ ✔

$\forall R_1, R_2 \in SO(3),\ R_1 + R_2 \in SO(3)$ ✘



- New tools – Lie group and Lie algebra
  - closed under addition and multiplication operations → Lie group
  - derivable → smooth and continuous → Lie group (differential manifold)
  - derivative operation → tangent space of Lie group → Lie algebra
  - non-linear optimization → Lie algebra attached to the vector space

Our method

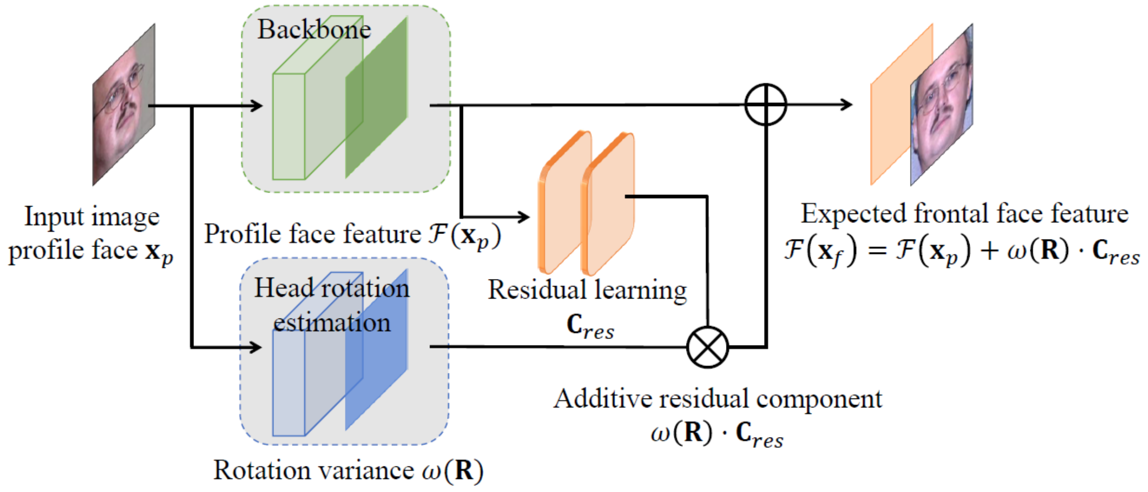
- The frontal-profile face feature pairs in ResNet based on the Lie algebra theory:

$$\mathcal{F}(\mathbf{x}_f) \approx \mathcal{F}(\mathbf{x}_p) + \mathcal{R}_{map}^{-1}(\mathcal{F}(\mathbf{x}_p) - \mathcal{R}_{map}\mathcal{F}(\mathbf{x}_p))$$

$$\mathcal{F}(\mathbf{x}_f) = \mathcal{F}(\mathbf{x}_p) + \omega(\mathbf{R}) \cdot \mathbf{C}_{res}(\mathbf{R}, \mathbf{x}_p)$$

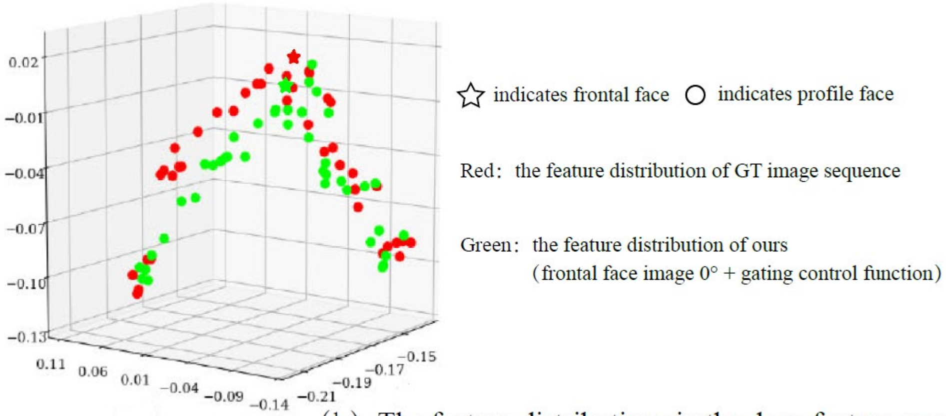
  - $\mathbf{C}_{res}$: a residual subnet for decoding rotation information from input face images
  - $\omega(\mathbf{R})$: a gating subnet to learn rotation magnitude for controlling the strength of $\mathbf{C}_{res}$

- Our succinct architecture:

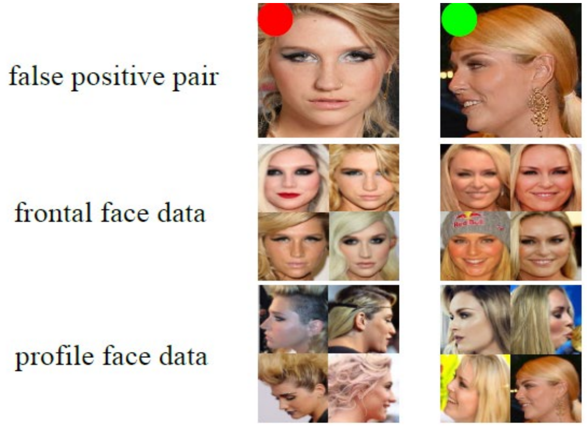


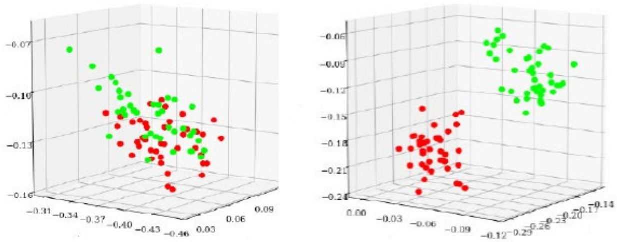
Visualization results – feature distributions

(a) A ground-truth image sequnce: pose varirant is from -90° to 90°.



(b) The feature distributions in the deep feature space





The feature distributions of the other method (left) and ours (right)

- ☐ Feature Representation

different-pose samples with the same identity

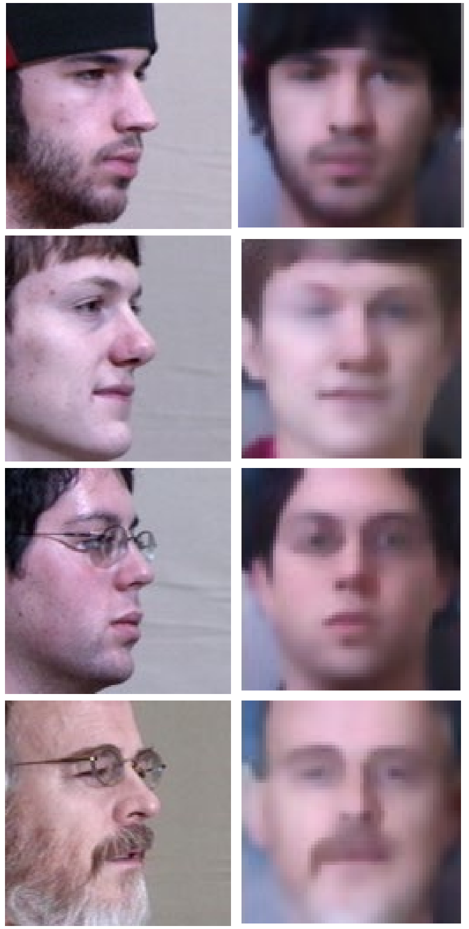
- ☐ Feature Clustering

different-pose samples with the different identities

Visualization results – reconstructed faces

15° profile faces

45° profile faces



60° profile faces

Quantitative evaluation

Method	TAR@FAR=0.01	TAR@FAR=0.001	Rank-1 Acc.	Rank-5 Acc.
Wang *et al.* (2016)	0.729	0.510	0.822	0.931
Pooling Faces (Hassner et al., 2016)	0.819	0.631	0.846	0.933
Multi Pose-Aware (AbdAlmageed et al., 2016)	0.787	—	0.846	0.927
DCNN Fusion (f.) (Chen et al., 2016)	0.838	—	0.903	0.965
PAMs (Masi et al., 2016a)	0.826	0.652	0.840	0.925
Augmentation+Rendered (Masi et al., 2016b)	0.886	0.725	0.906	0.962
Multi-task learning (Yin & Liu, 2017)	0.787	—	0.858	0.938
TPE(f.) (Sankaranarayanan et al., 2017)	0.900	0.813	0.932	—
DR-GAN (Tran et al., 2017)	0.831	0.699	0.901	0.953
FF-GAN (Yin et al., 2017)	0.852	0.663	0.902	0.954
NAN (Yang et al., 2017)	0.921	0.861	0.938	0.960
Multicolumn (Xie & Zisserma, 2018)	0.920	—	—	—
VGGFace2 (Cao et al., 2018b)	0.904	—	—	—
Template Adaptation(f.) (Crosswhite et al., 2018)	0.939	—	0.928	—
DREAM (Cao et al., 2018a)	0.872	0.712	0.915	0.962
DREAM(E2E+retrain,f.) (Cao et al., 2018a)	0.934	0.836	0.939	0.960
FTL with 60K parameters (o.s.) (Yin et al., 2019)	0.864	0.744	0.893	0.947
PFEs (Shi & Jain, 2019)	0.944	—	—	—
DebFace (Gong et al., 2020)	0.902	—	—	—
Rotate-and-Render (Zhou et al., 2020)	0.920	0.825	—	—
HPDA (Wang et al., 2020)	0.876	0.803	0.84	0.88
CDA (Wang & Deng, 2020)	0.911	0.823	0.936	0.957
LARNet	0.941	0.842	0.936	0.968
LARNet+	**0.951**	**0.874**	**0.949**	**0.971**

- Quantitative evaluation on IJB-A dataset

Our results surpass many competitors in the recent 5 years under all 4 evaluation criteria:

- Verification task

  achieve 95%+ under TAR@FAR=0.01

  & 87%+ under TAR@FAR=0.001

- Identification task

  achieve 95% under Rank-1 recognition accuracy

  & 97%+ under Rank-5 recognition accuracy

Quantitative evaluation

- Quantitative evaluation on CFP-FP dataset

Method	Verification (%)
SphereFace (o.s.+f.)	94.17
CosFace (o.s.)	94.40
ArcFace (o.s.+f.)	94.04
URFace (all modules, MS1MV2, o.s.)	98.64
Human-level	98.92
LARNet	98.84
LARNet+	**99.21**

- Quantitative evaluation on LFW,YTF, and CPLFW datasets

Method	LFW	YTF	CPLFW
HUMAN-Individual	97.27	-	81.21
HUMAN-Fusion	**99.85**	-	85.24
DeepID (Sun et al., 2014)	99.47	93.20	-
Deep Face (Taigman et al., 2014)	97.35	91.4	-
VGG Face (Parkhi et al., 2015)	98.95	97.30	90.57
FaceNet (Schroff et al., 2015)	99.63	95.10	-
Baidu (Liu et al., 2015a)	99.13	–	-
Center Loss (Wen et al., 2016)	99.28	94.9	85.48
Range Loss (Zhang et al., 2017)	99.52	93.70	-
Marginal Loss (Deng et al., 2017)	99.48	95.98	-
SphereFace(o.s.) (Liu et al., 2017)	99.42	95.0	81.4
SphereFace+(o.s.)	99.47	–	90.30
CosFace(o.s.) (Wang et al., 2018a)	99.51	96.1	-
CosFace*(MS1MV2,R64, o.s.)	99.73	97.6	-
Arcface(o.s.) (Deng et al., 2019)	99.53	–	92.08
ArcFace*(MS1MV2,R100,f.)	99.83	**98.02**	95.45
Ours: LARNet	99.36	96.55	95.51
Ours: LARNet+	99.71	97.63	**96.23**

- The first result achieves 99%+ on CFP-FP dataset.
- The first result surpasses the reported human-level.
- Our method also outperforms on general face recognition datasets.

THANKS

LARNet: Lie Algebra Residual Network for Face Recognition

yangxiaolong17@mails.ucas.ac.cn

paradocx.github.io

Paper Homepage

QR Code