JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Pose-disentangled Contrastive Learning for Self-supervised Facial Representation

Yuanyuan Liu, Wenbin Wang, Yibing Zhan, Shaoze Feng, Kejun Liu, Zhe Chen

Poster: WED-AM-142

Paper: https://arxiv.org/abs/2211.13490

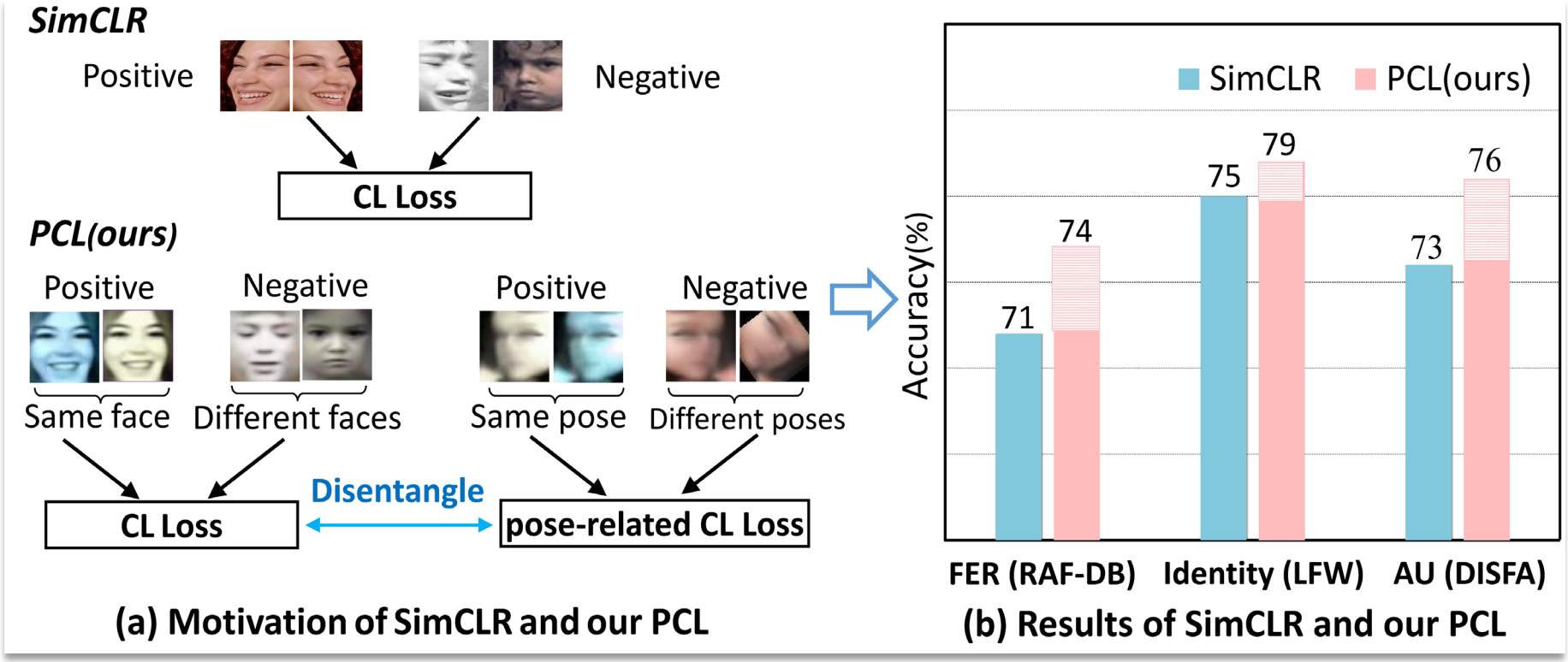
Code: https://github.com/DreamMr/PCL

中国地质大学
CHINA UNIVERSITY OF GEOSCIENCES
武汉 · WUHAN

THE UNIVERSITY OF SYDNEY

JD.京东.COM

PCL overview

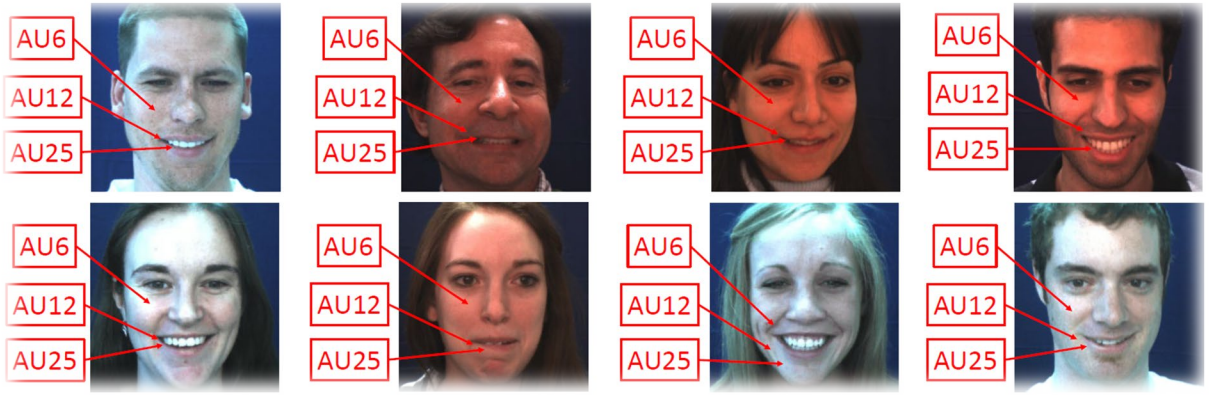


- PCL improve the SimCLR for general facial repersentation learning.
- PCL introduces pose-disentangled decoder (PDD) and pose-related contrastive learning scheme.
- PCL achieves SOTA on four challenging downstream facial understanding tasks.

🤔 Motivation

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

FER dataset

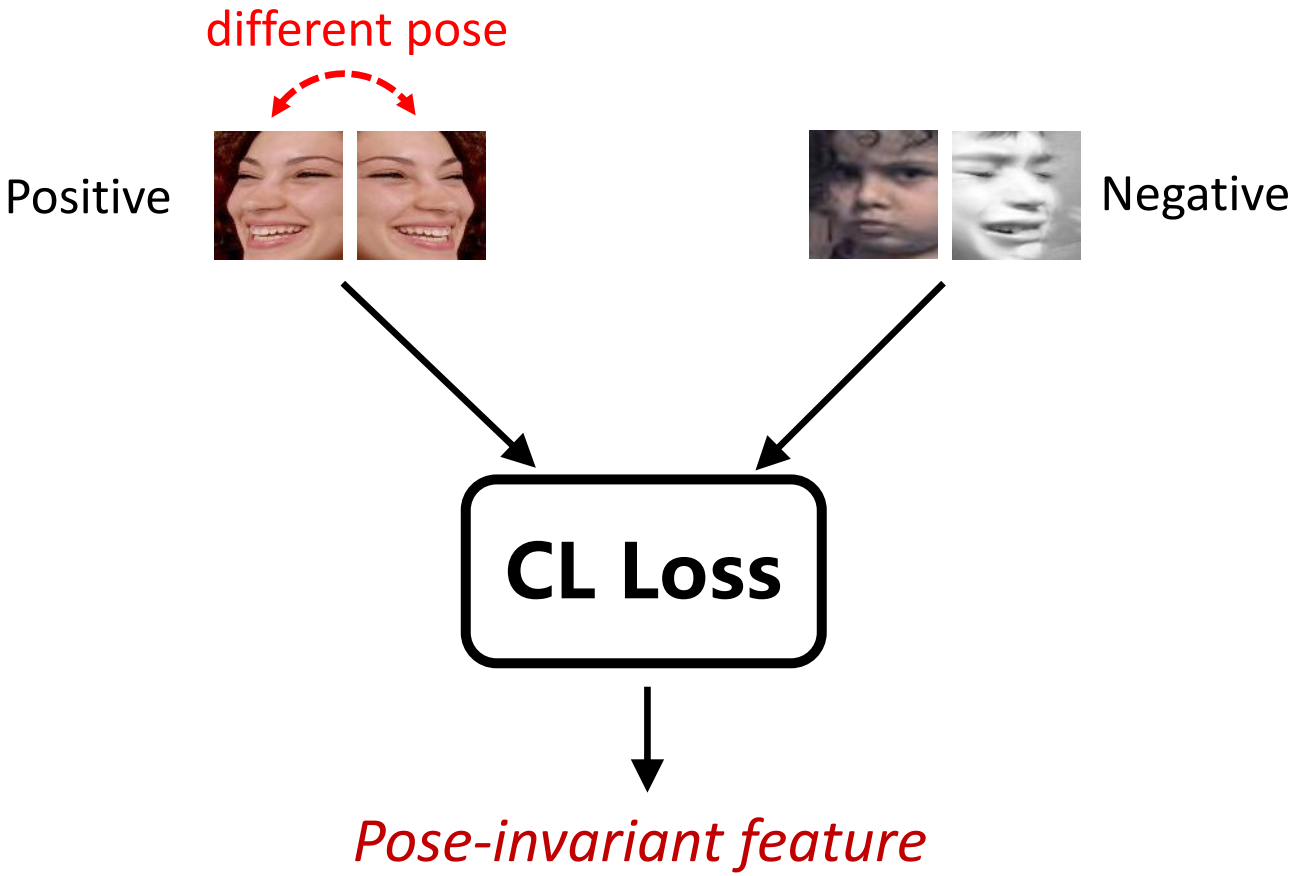


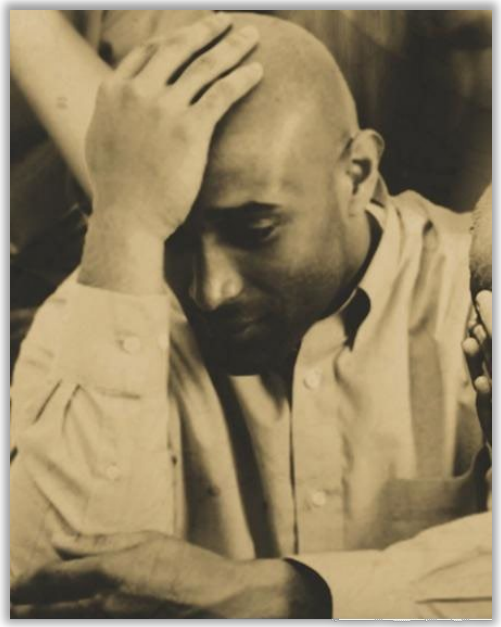
AU detection dataset

Labeling face data is a time-costly process

How to use unlabeled facial data to learn general facial representation ?

Motivation



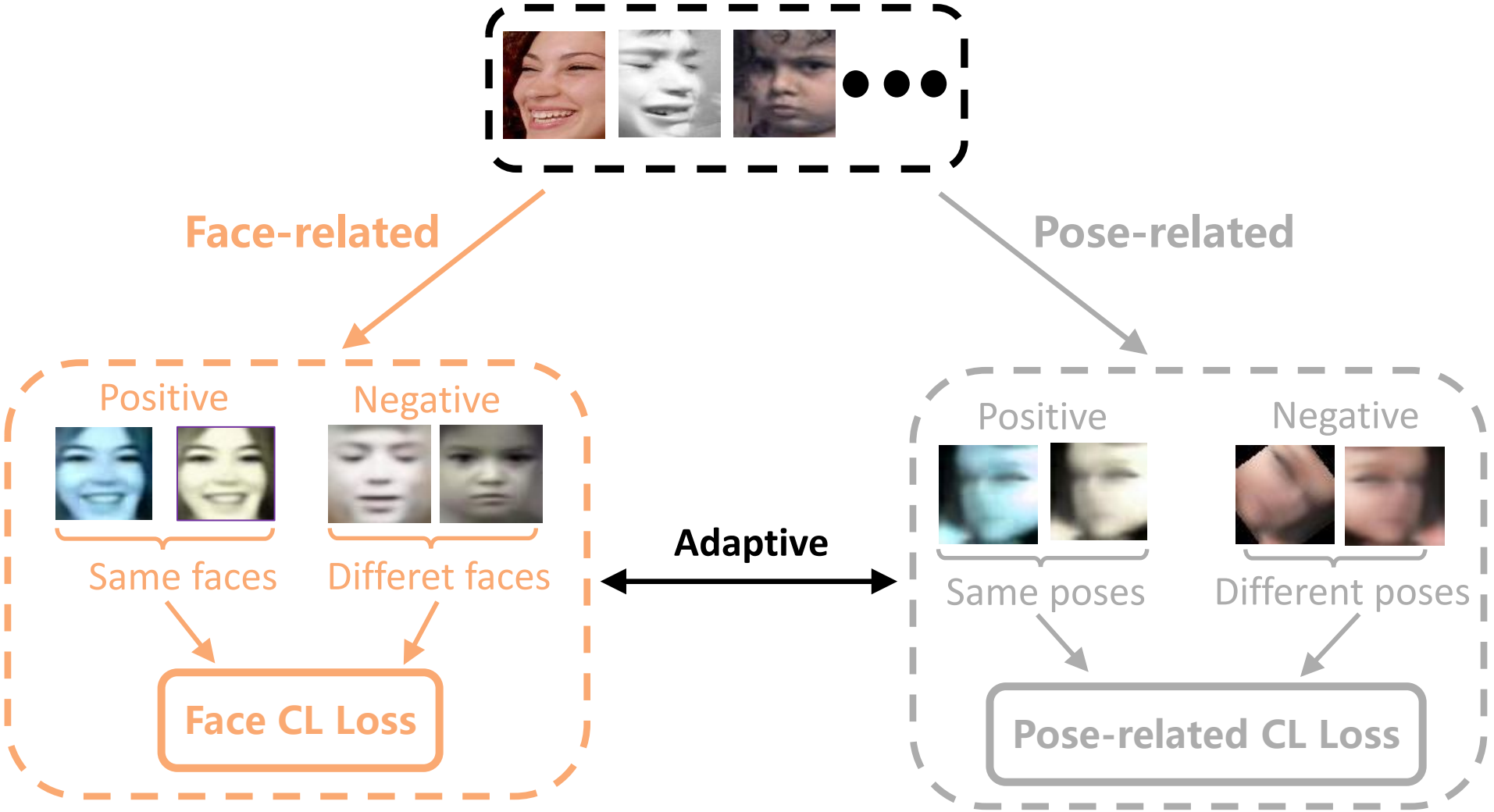


i.e., a person tends low their head when they feel sad

Poses are one significant consideration for facial understanding

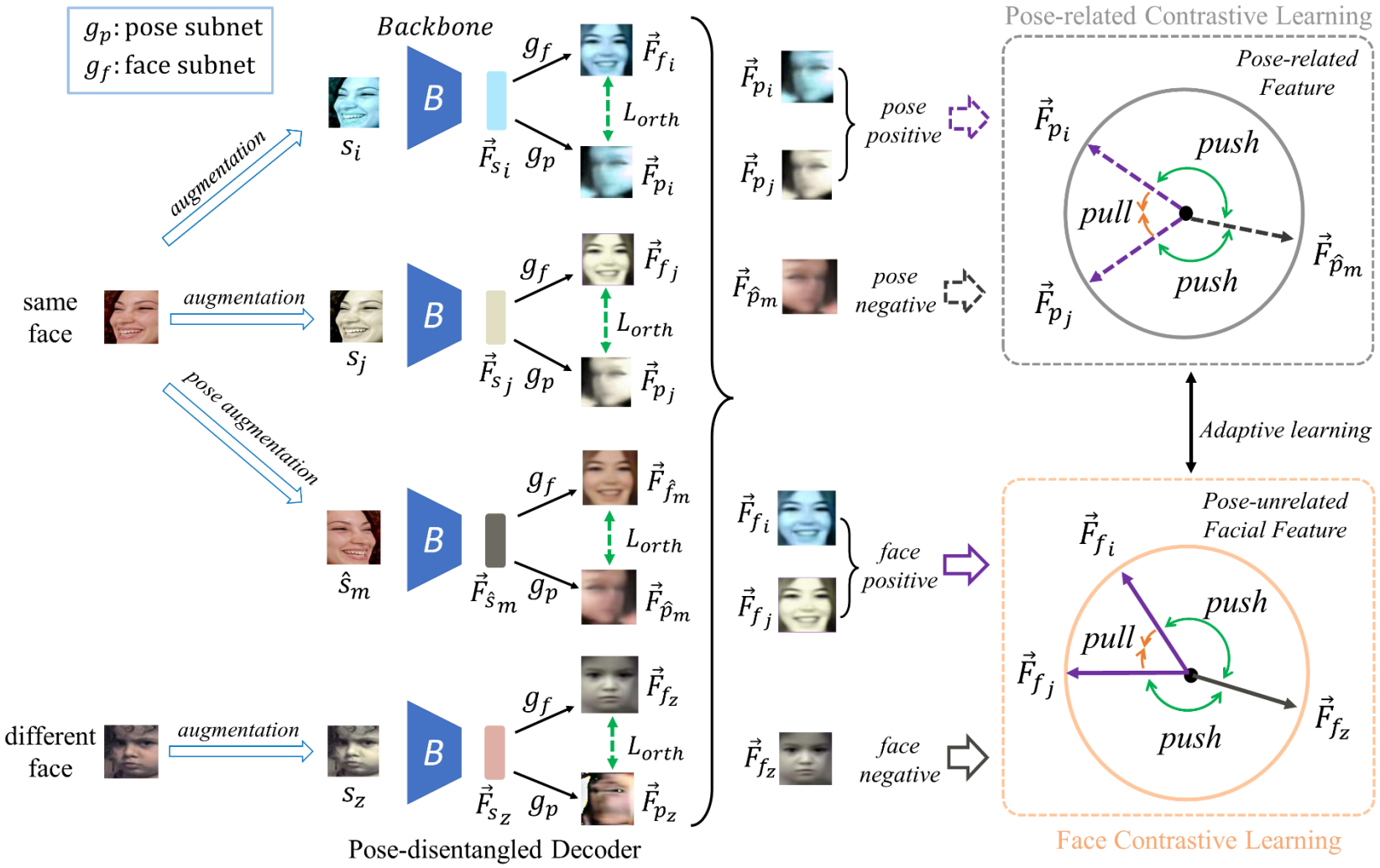
Motivation

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



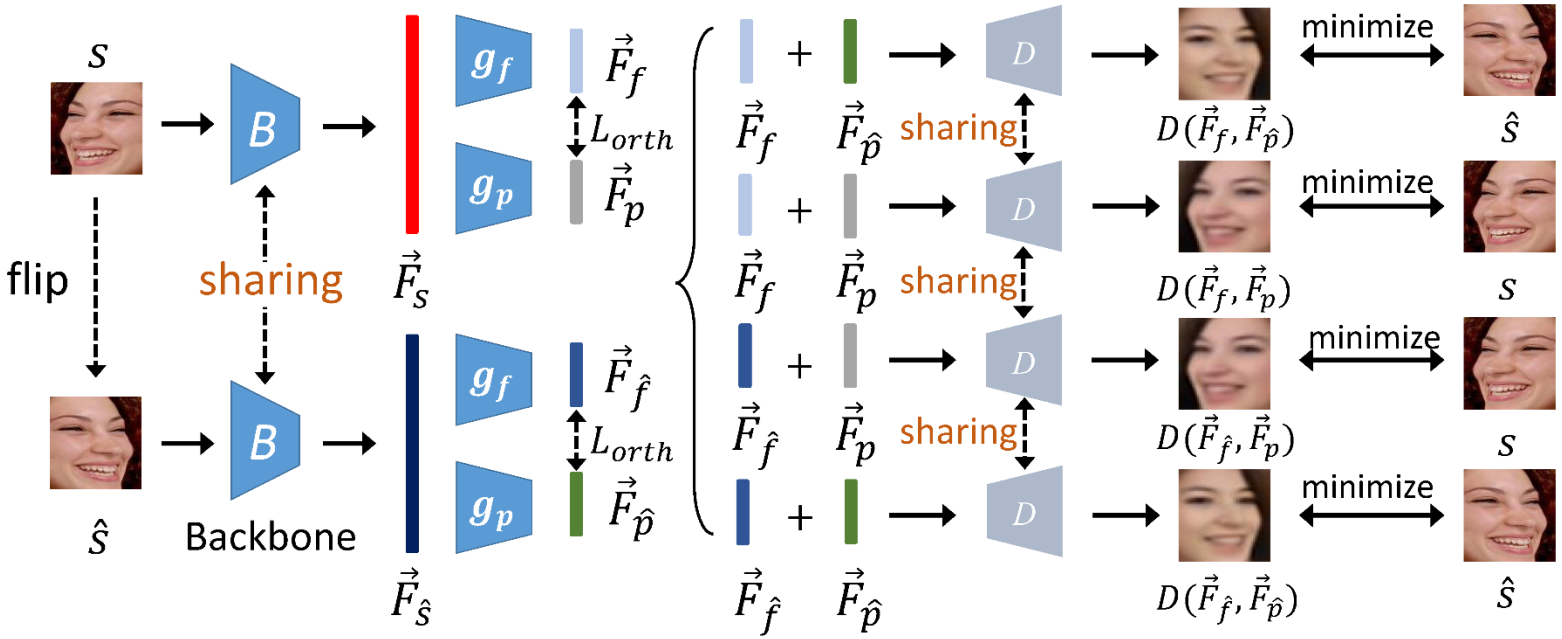
Method

- **Framework**



Method

- **Pose-disentangled decoder (PDD)**

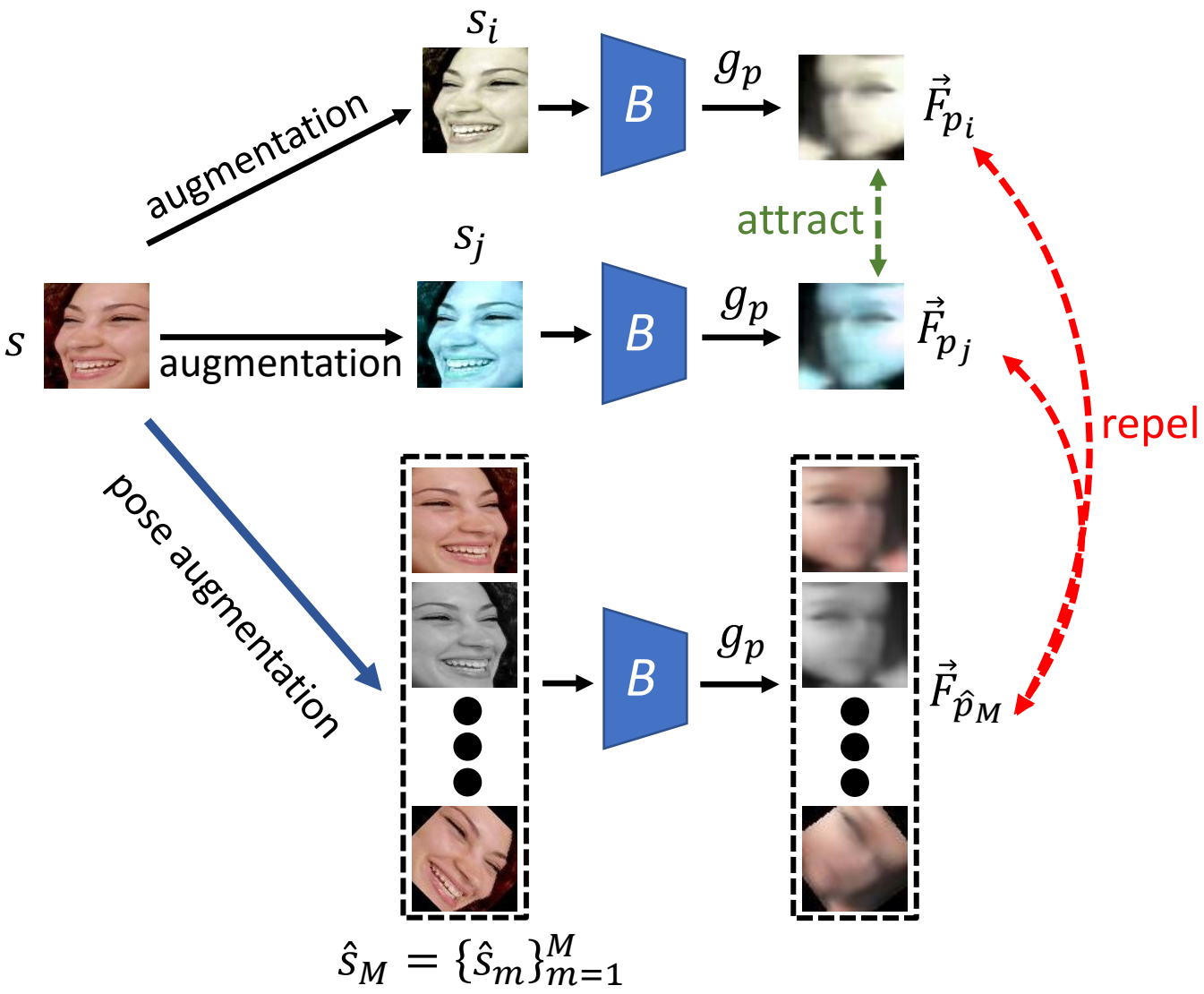


$$\mathcal{L}_{dis} = ||s - D(\vec{F}_f, \vec{F}_p)||_1 + ||\hat{s} - D(\vec{F}_f, \vec{F}_{\hat{p}})||_1 + ||s - D\left(\vec{F}_{\hat{f}}, \vec{F}_p\right)||_1 + ||\hat{s} - D\left(\vec{F}_{\hat{f}}, \vec{F}_{\hat{p}}\right)||_1$$

$$\mathcal{L}_{orth} = \frac{1}{N}(\sum_{i=1}^{N} ||\vec{F}_f \cdot \vec{F}_p||_2^2 + ||\vec{F}_{\hat{f}} \cdot \vec{F}_{\hat{p}}||_2^2)$$

Method

- Pose-related Contrastive Learning

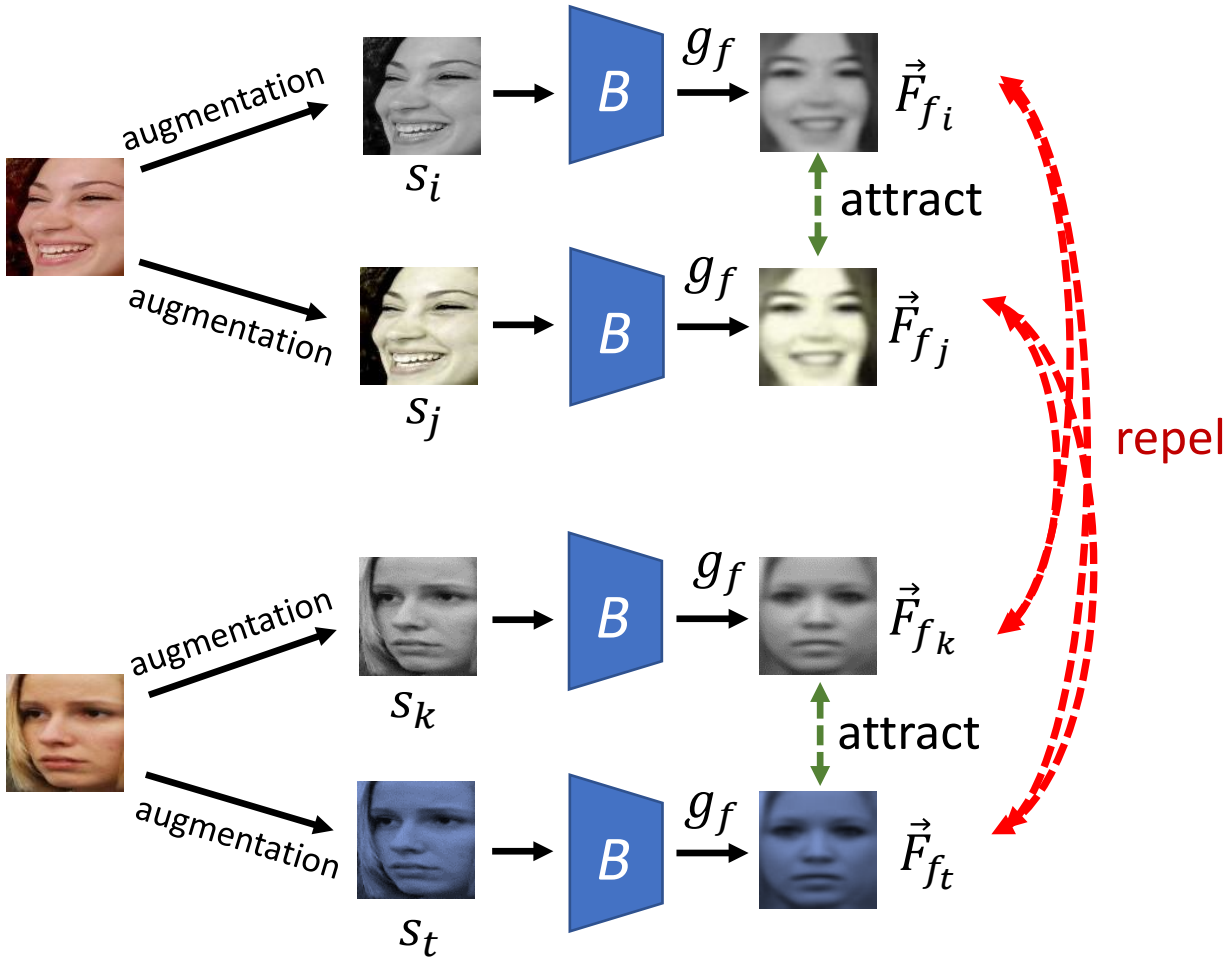


$$l_p\left(\vec{F}_{p_i}, \vec{F}_{p_j}\right) = -log\frac{\exp(\frac{sim(\vec{F}_{p_i}, \vec{F}_{p_j})}{\tau})}{\sum_{m=1}^{M}\exp(\frac{sim(\vec{F}_{p_i}, \vec{F}_{\hat{p}_m})}{\tau})}$$

$$\mathcal{L}_{pose}\left(\vec{F}_{p_i}, \vec{F}_{p_j}, \vec{F}_{\hat{p}_M}\right) = l_p\left(\vec{F}_{p_i}, \vec{F}_{p_j}\right) + l_p(\vec{F}_{p_j}, \vec{F}_{p_i})$$

Method

- Face-related Contrastive Learning

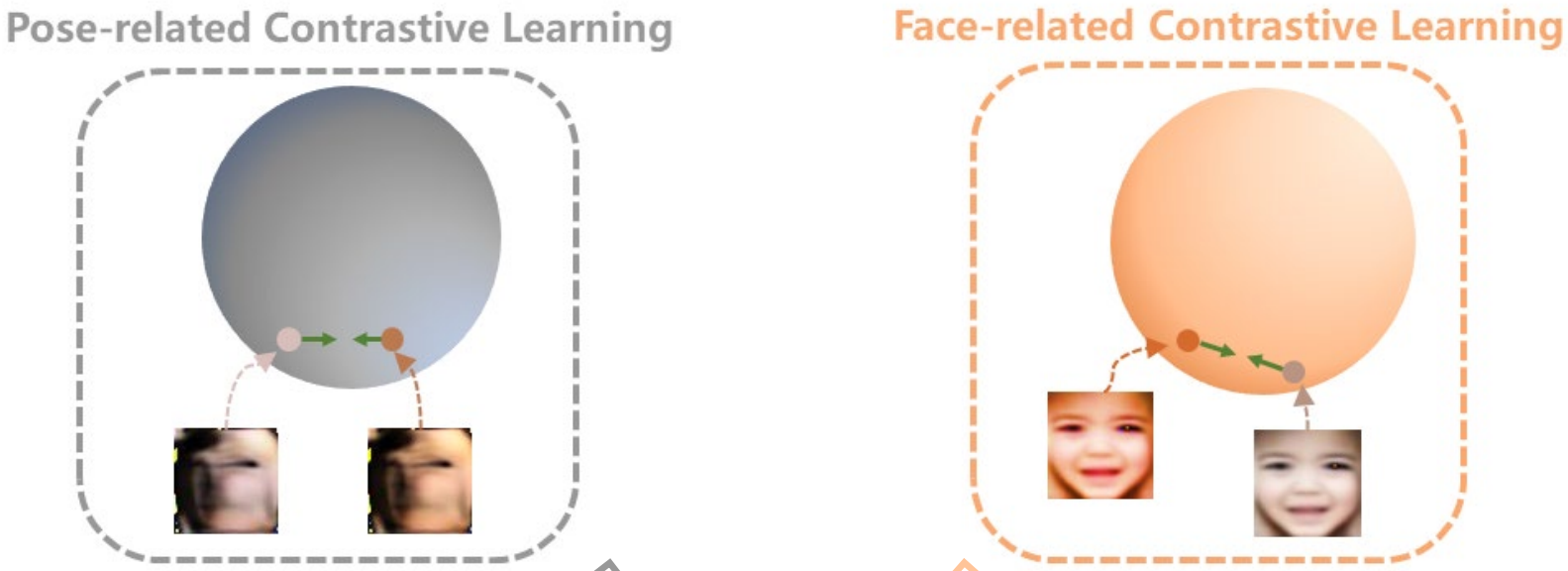


$$\mathcal{L}_{face}\left(\vec{F}_{f_a}, \vec{F}_{f_b}\right) = l_f\left(\vec{F}_{f_a}, \vec{F}_{f_b}\right) + l_f(\vec{F}_{f_a}, \vec{F}_{f_b})$$

$$l_f\left(\vec{F}_{f_a}, \vec{F}_{f_b}\right) = -log\frac{\exp(\frac{sim(\vec{F}_{f_a}, \vec{F}_{f_b})}{\tau})}{\sum_{z=1}^{2N} 1_{[i\neq z]}\exp(\frac{sim(\vec{F}_{f_a}, \vec{F}_{f_z})}{\tau})}$$

Method

- **Overall Optimization Objectives**



$$\mathcal{L} = \mathcal{L}_{PDD} + \underbrace{\alpha_{pose} \cdot \mathcal{L}_{pose} + \alpha_{face} \cdot \mathcal{L}_{face}}_{\text{Dynamic Weight Average}}$$

Experiments

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Facial Expression Recognition

	FER-2013	RAF-DB
Method	Accuracy(%)	Accuracy(%)
Fully supervised		
FSN	67.60	81.10
ALT	69.85	84.50
Self-supervised (linear evaluation)		
LBP	37.89	52.17
HoG	45.47	63.53
FAb-Net	46.98	66.72
TCAE	45.05	65.32
MoCo	47.24	68.32
FaceCycle	48.76	71.01
SimCLR	49.51	71.06
Ours	**56.81**	**74.47**

Face Recognition

	LFW	CPLFW
Method	Accuracy(%)	Accuracy(%)
Fully supervised		
VGG-Face	98.95	84.00
SphereFace	99.42	81.40
ArcFace	99.53	92.08
Self-supervised (linear evaluation)		
LBP	72.44	-
VGG	72.20	-
MoCo	65.88	57.82
FaceCycle	74.12	63.35
SimCLR	75.97	62.25
Ours	**79.72**	**64.61**

Experiments

- **Evaluation for Head Pose Regression**

	AFLW2000 (trained on 300W-LP)				BU-3DFE
Method	Yaw↓	Pitch↓	Roll↓	MAE↓	Accuracy(%)↑
Fully supervised					
FAN	6.36	12.3	8.71	9.12	-
FSA-Net	5.27	6.71	5.28	5.75	-
Self-supervised (linear evaluation)					
Dlib(68 points)	23.10	13.60	**10.50**	15.80	-
MoCo	28.49	16.29	15.55	20.11	75.33
FaceCycle	11.70	**12.76**	12.94	12.47	98.82
SimCLR	11.20	19.86	12.08	14.38	98.85
Ours	**9.86**	16.59	10.62	**12.36**	**98.95**

Experiments

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- **Evaluation for Facial AU Detection**

	Methods/AU	1	2	4	6	9	12	25	26	ave
Supervised	DRML	17.3	17.7	37.4	29.0	10.7	37.7	38.5	20.1	26.7
	EAC-Net	41.5	26.4	66.4	50.7	80.5	89.3	88.9	15.6	48.5
	JAA-Net	43.7	46.2	56.0	41.4	44.7	69.6	88.3	58.4	56.0
Self-supervised	SplitBrain	13.1	10.6	35.7	40.2	30.2	57.5	77.4	40.3	38.1
	DeformAE	17.6	12.3	46.7	43.5	26.0	62.7	64.8	**47.6**	40.1
	Fab-Net	15.5	16.2	43.2	**50.4**	23.2	69.6	72.4	42.4	41.6
	TCAE	15.1	16.2	50.5	48.7	23.3	72.1	72.4	42.4	45.0
	FaceCycle	26.4	10.2	37.3	21.5	25.0	71.8	84.2	34.7	38.9
	SimCLR	40.5	**46.9**	53.8	33.5	24.9	74.7	85.0	35.6	49.4
	Ours	**53.8**	44.9	**58.1**	37.2	**53.2**	**73.1**	**86.5**	31.1	**54.8**

Experiments

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Ablation study

Baseline(SimCLR)	PDD		Contrastive learning		FER
	$\mathcal{L}_{dis}$	$\mathcal{L}_{orth}$	$\mathcal{L}_{pose}$	Dynamic weighting	
✓					71.06
✓	✓				71.47
✓	✓	✓			72.39
✓	✓	✓	✓		73.73
✓	✓	✓	✓	✓	**74.47**

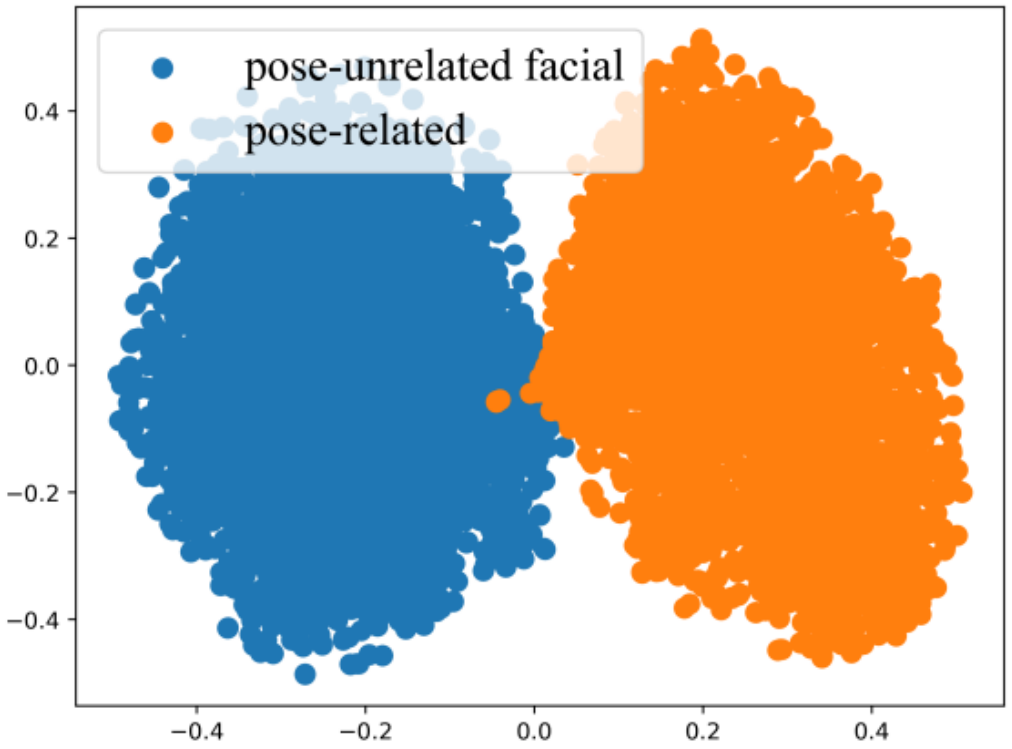
Comparison of Different Learned Features

Different features	RAF-DB	LFW	DISFA
$\vec{F}_f$	73.04	78.55	54.30
$\vec{F}_f$	65.71	62.55	34.17
$\vec{F}_s$	**74.47**	**79.72**	54.78
$\vec{F}_f + \vec{F}_p$	73.53	79.10	**56.26**

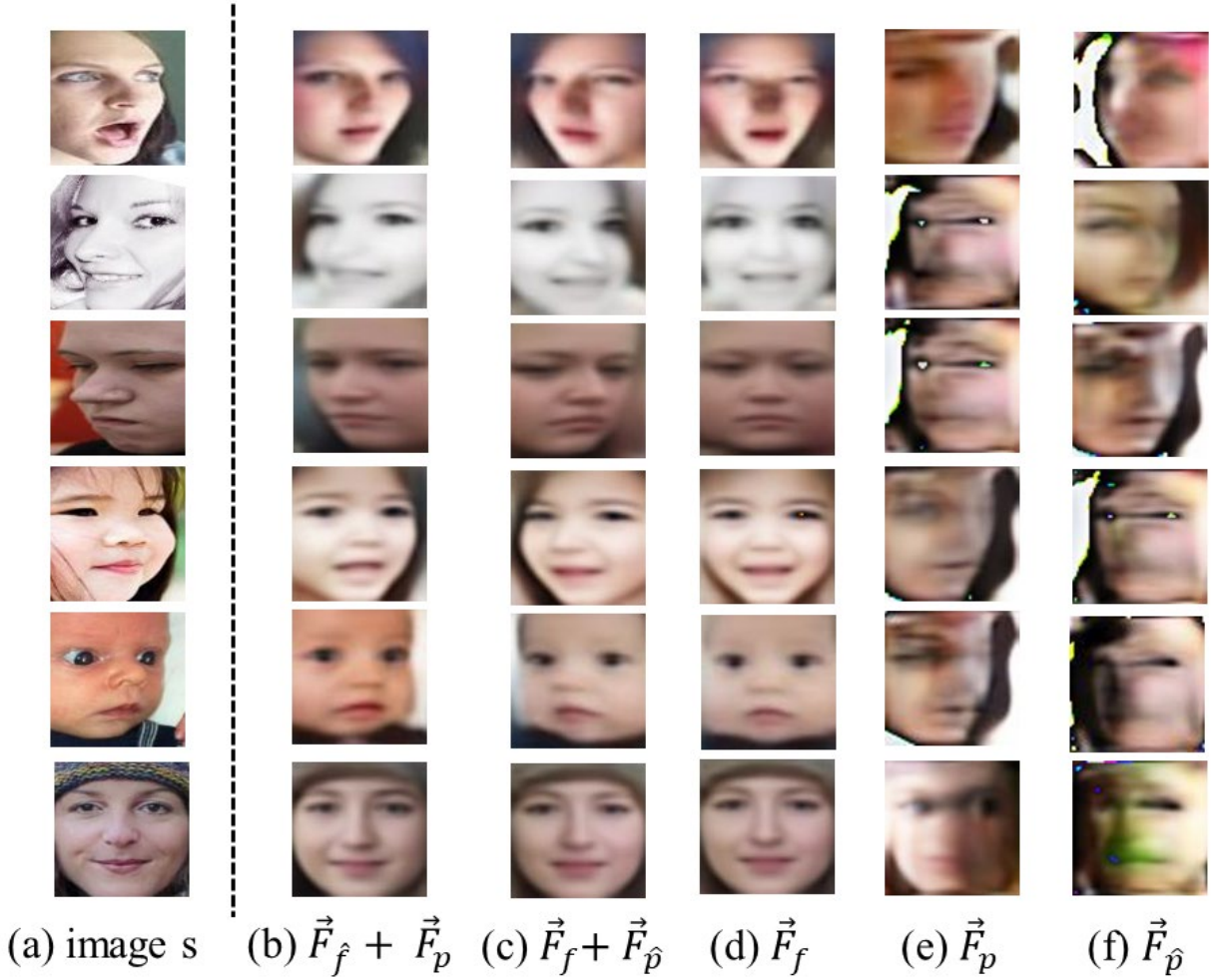
Experiments

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- The disentangled pose-related and pose-unrelated facial features

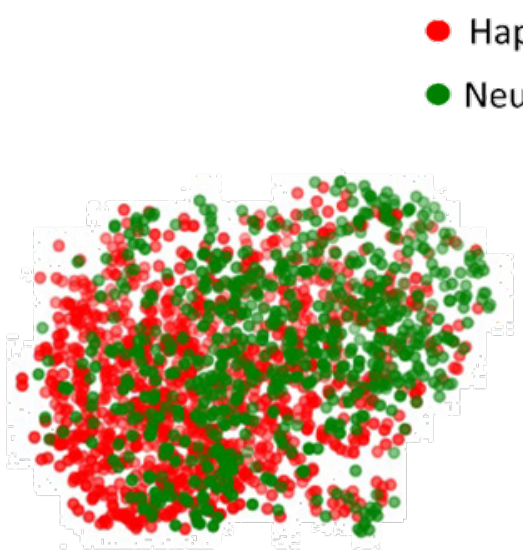


- The reconstructed faces with the disentangled features.

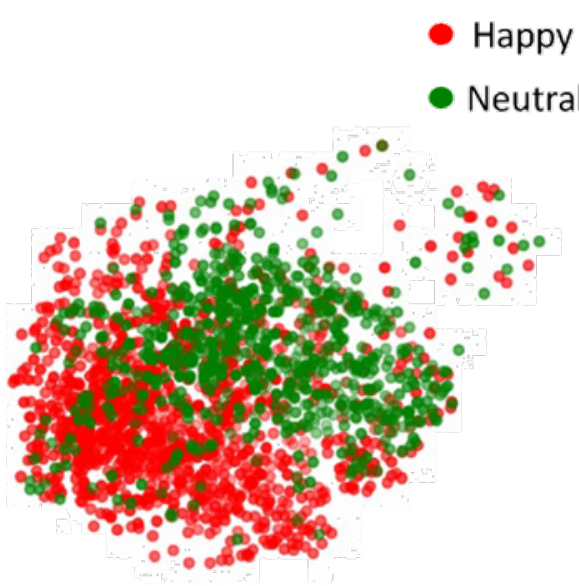


Experiments

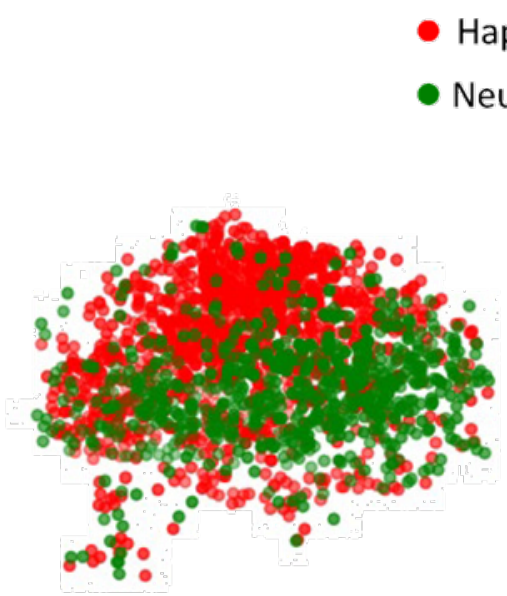
- **The features learned by SimCLR and our PCL respectively visualized by t-SNE. Our PCL can learn better distinguishability features.**



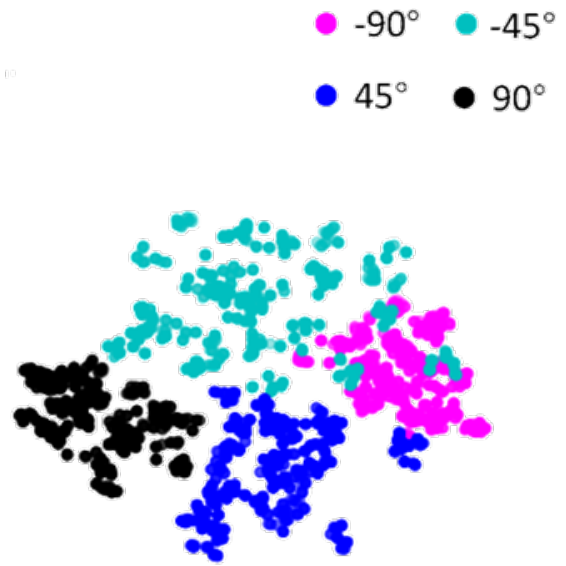
(a) SimCLR



(b) $\vec{F}_s$



(c) $\vec{F}_f$



(d) $\vec{F}_p$

Conclusions

- The PCL is a novel pose-disentangled contrastive learning framework.
- The PCL design a pose-disentangled decoder to separate the face-aware features obtained from the backbone into pose-related and pose-unrelated facial features.
- The PCL introduce a pose-related contrastive learning scheme for pose-related feature learning.
- The PCL can be well generalized to several downstream tasks, e.g., facial expression recognition, facial AU detection, facial recognition, and head pose estimation.

👏 Thank you !