Personalized Feature Translation for Expression Recognition: An Efficient Source-Free Domain Adaptation Method

Masoumeh Sharafi[1], Soufiane Belharbi[1], Muhammad Osama Zeeshan[1], Houssem Ben Salem[1], Ali Etemad[3], Alessandro Lameiras Koerich[1], Marco Pedersoli[1], Simon L Bacon[2], Eric Granger[1]

[1] LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada
[2] Dept. of Health, Kinesiology & Applied Physiology, Concordia University, Montreal, Canada
[3] Dept. of Electrical and Computer Engineering, Queen's University, Kingston, Canada

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
Le génie pour l'industrie
Université du Québec

LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE

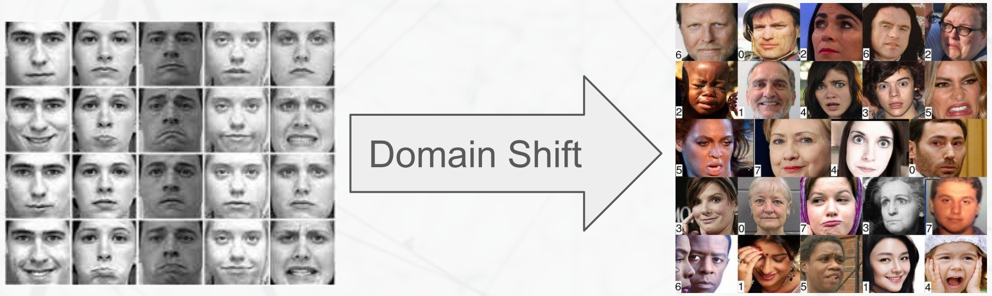
CHAIRE DOUBLE FRSQ EN IA ET EN SANTÉ NUMÉRIQUE

FRSQ DOUBLE CHAIR IN AI AND DIGITAL HEALTH

Outline

Context: Facial Expression Recognition in Video

Train Test



Source Target

Challenges:

- Subtle expression in health applications
- Variability of dynamic expressions captured in videos
- Variability across individuals and capture conditions (sensors, devices,
- and environments)
- Real-world capture: occlusion, pose, illumination, complex background, noise
- Shift in distributions between the design (source) and operational (target) datasets
- Cost of capturing and annotating large-scale datasets that are realistic

Variation in capture conditions

Subtle expressions: frames with pain (left) vs. no pain (right)

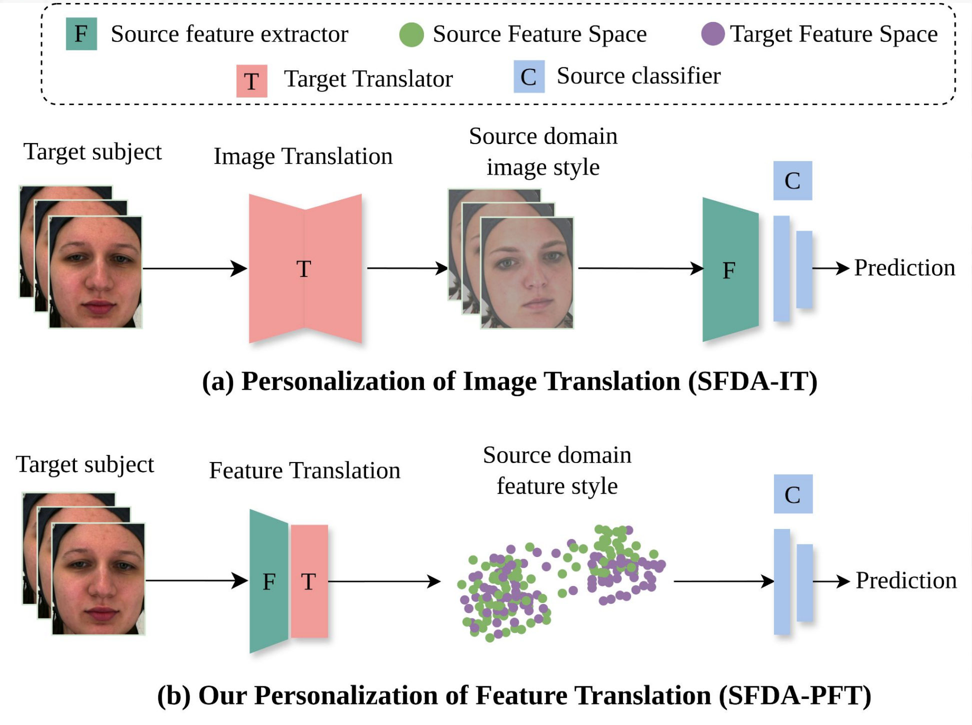
Li, S., Deng, W. and Du, J., Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. CVPR 2017.

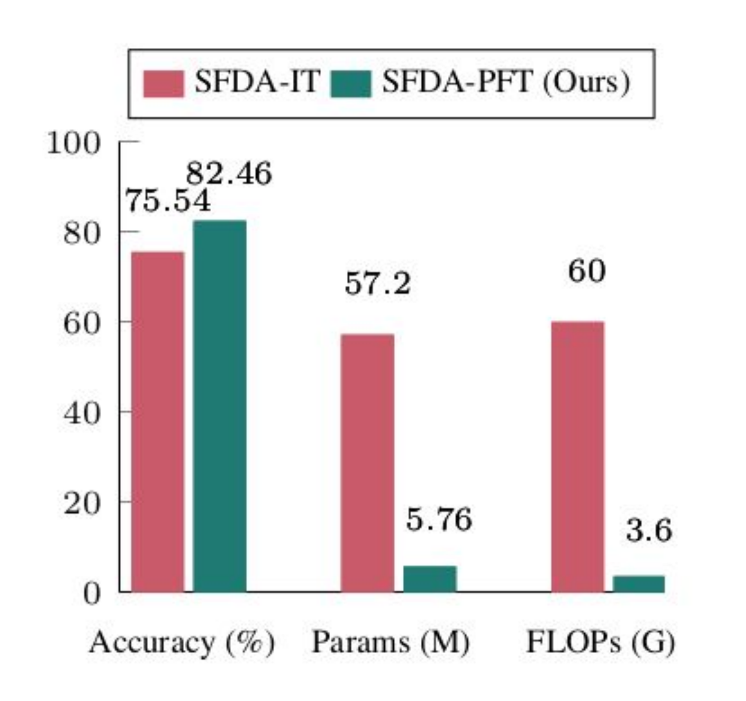
Motivation: Personalization of Models for Expression Recognition

- **Personalization:** Deep FER models often degrade on new users due to domain shift and strong inter-subject variability in facial expressions.
- **Source-free domain adaptation (SFDA):** It provides a practical framework by adapting a pretrained FER model using only unlabeled target video data without access to the source data.
- **Limited-classes:** SFDA methods assume target data contains diverse expression classes, whereas real-world personalization data is often limited to short neutral-only videos due to privacy concerns.

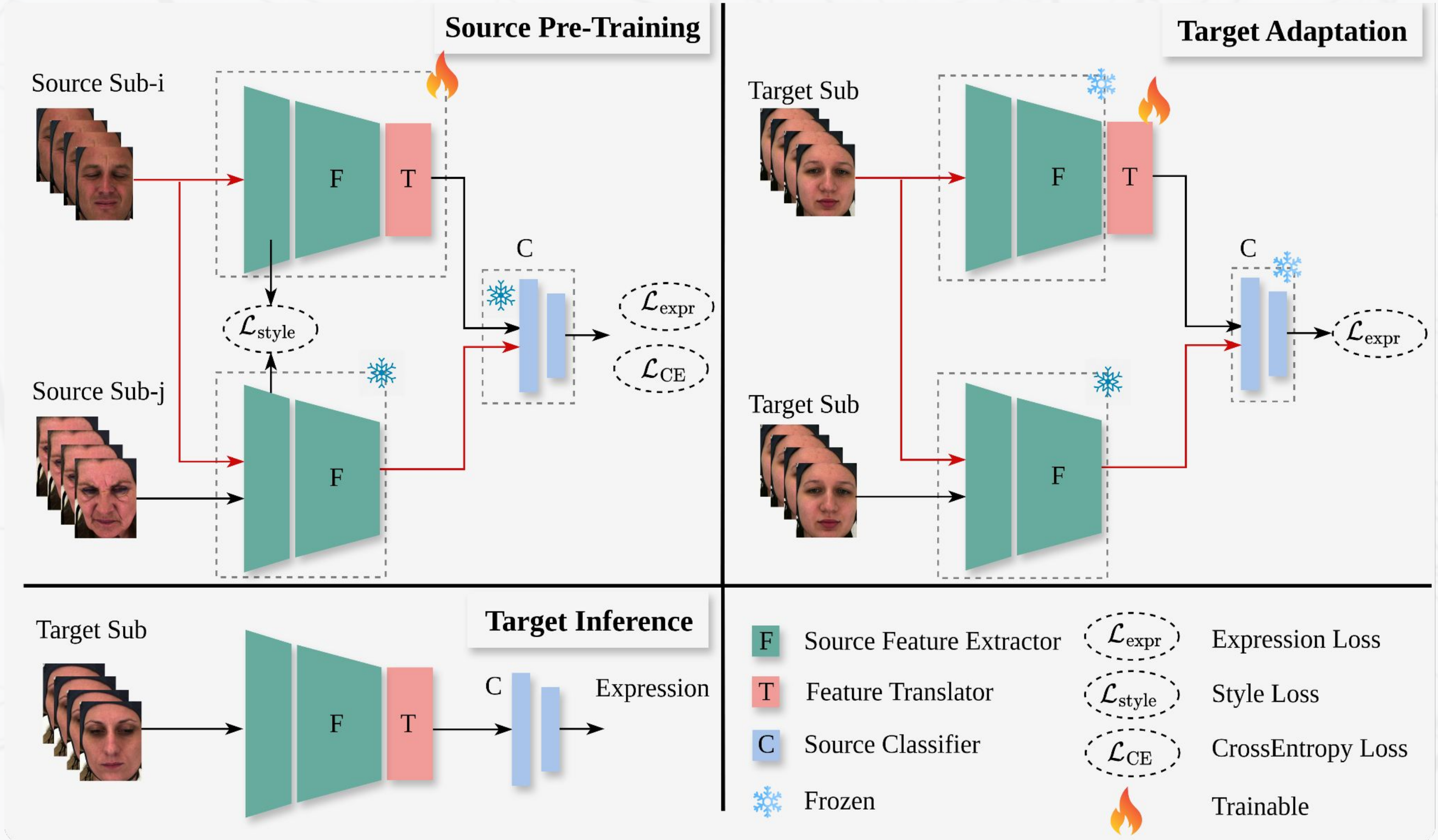
Main Contributions

- SFDA with personalized feature translation (SFDA-PFT) uses only neutral target data by translating in feature space instead of pixel space with a lightweight model
- Style-aware and expression consistency losses preserve face expression semantics during cross-subject feature translation
- Higher performance on four FER benchmarks with lower complexity, no image synthesis, and no extra inference-time parameters





Proposed SFDA-PFT Method



Source pre-training: train a translator T to map features across source subjects while preserving expression semantics

Target adaptation: adapt T using unlabeled target videos through prediction consistency between two target facial images, while keeping C fixed

Target Inference: use adapted T and fixed classifier C to predict target expressions

Identity alignment loss

$$\mathcal{L}_{\text{style}} = \sum_{l \in \mathcal{L}} \left(\|\mu(\hat{\mathbf{f}}_i^l) - \mu(\mathbf{f}_j^l)\|_2^2 + \|\sigma(\hat{\mathbf{f}}_i^l) - \sigma(\mathbf{f}_j^l)\|_2^2 \right)$$

Expression preservation

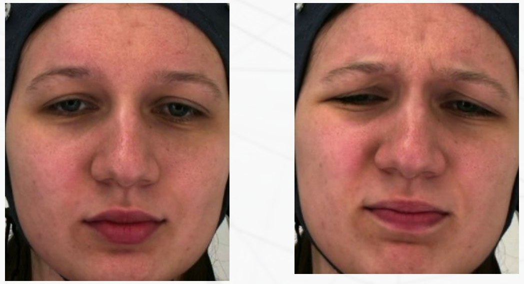
$$\mathcal{L}_{\text{expr}} = D_{\text{KL}}\left(\mathbf{C}(\mathbf{f}_i) \,\|\, \mathbf{C}(\hat{\mathbf{f}}_i)\right)$$

Source classification loss

$$\mathcal{L}_{\text{CE}}(\mathbf{x}_s, y_s) = -\log\left[\mathbf{C}(\mathbf{F}(\mathbf{x}_s))\right]_{y_s}$$

Results: Datasets and Settings

BioVid Pain and Heat Dataset



neutral pain

StressID Dataset

neutral stress

Behavioural Ambivalence/Hesitancy Dataset

neutral ambivalence/ hesitancy

Aff-wild2 Dataset

neutral anger disgust fear happiness surprise

Settings

- Source-only
- SFDA (model-based)
- SFDA (data-based)
- Oracle

Walter, S., Gruss, S., Ehleiter, H., Tan, J., Traue, H. C., Werner, P., ... & Da Silva, G. M. (2013, June). The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In *2013 IEEE international conference on cybernetics (CYBCO)* (pp. 128-131). IEEE.

Chaptoukaev, H., Strizhkova, V., Panariello, M., Dalpaos, B., Reka, A., Manera, V., ... & M Ferrari, L. (2023). Stressid: a multimodal dataset for stress identification. Advances in Neural Information Processing Systems, 36, 29798-29811.

González-González, M., Belharbi, S., Zeeshan, M. O., Sharafi, M., Aslam, M. H., Pedersoli, M., ... & Granger, E. (2025). Bah dataset for ambivalence/hesitancy recognition in videos for behavioural change. arXiv preprint arXiv:2505.19328.

Kollias, D., & Zafeiriou, S. (2018). Aff-wild2: Extending the aff-wild database for affect recognition. arXiv preprint arXiv:1811.07770.

Comparison with State-of-the-Art Methods

BioVid dataset (10 target subjects, 77 source subjects). Bold numbers indicate the best F1 score.

Setting	Methods	Sub-1	Sub-2	Sub-3	Sub-4	Sub-5	Sub-6	Sub-7	Sub-8	Sub-9	Sub-10	Average
Source-only	Source model (no adaptation)	62.78	52.76	82.02	80.83	82.73	56.03	71.85	66.90	50.01	45.79	65.17
SFDA (model-based)	SHOT (Liang et al., 2020)	52.97	45.35	38.98	49.80	51.92	46.43	51.72	46.74	52.10	42.20	47.82
	NRC (Yang et al., 2021)	48.45	32.16	68.60	59.52	65.06	34.85	52.20	44.06	44.82	34.68	48.44
	TPDS (Tang et al., 2024)	62.26	53.16	75.23	64.79	87.06	56.14	58.20	65.84	54.24	45.79	62.27
	DSFDA (Sharafi et al., 2025)	65.72	64.10	77.57	73.12	75.20	57.59	76.15	74.73	59.08	61.54	68.48
SFDA (data-based)	SFIT (Hou & Zheng, 2021b)	76.85	65.33	78.70	80.44	87.01	54.44	57.54	70.81	57.66	**75.92**	70.47
	SFDA-IT (Hou & Zheng, 2021a)	71.54	63.89	84.53	80.30	86.24	59.18	77.66	72.08	54.97	67.01	71.74
	SFDA-PFT (ours)	**80.65**	**71.75**	**90.26**	**81.54**	**92.68**	**70.06**	**84.26**	**79.29**	**74.53**	58.08	**78.31**
Oracle	Supervised fine-tuning	92.22	86.83	91.89	92.96	91.27	87.65	85.48	90.30	93.28	92.12	90.40

StressID dataset (10 target subjects, 44 source subjects). Bold numbers indicate the best F1 score.

Setting	Methods	Sub-1	Sub-2	Sub-3	Sub-4	Sub-5	Sub-6	Sub-7	Sub-8	Sub-9	Sub-10	Average
Source-only	Source model (no adaptation)	44.44	43.54	45.34	44.89	45.79	43.99	45.34	44.89	44.44	45.34	44.80
SFDA (model-based)	SHOT (Liang et al., 2020)	42.66	41.79	43.52	43.09	43.95	42.22	43.52	43.09	42.66	43.52	43.00
	NRC (Yang et al., 2021)	40.67	39.85	41.49	41.08	41.90	40.26	41.49	41.08	40.67	41.49	41.00
	TPDS (Tang et al., 2024)	50.10	49.08	51.11	50.60	51.61	49.59	51.11	50.60	50.10	51.11	50.50
	DSFDA (Sharafi et al., 2025)	65.47	64.15	66.79	66.13	67.45	64.81	66.79	66.13	65.47	66.79	66.00
SFDA (data-based)	SFIT (Hou & Zheng, 2021b)	62.00	60.75	63.25	62.63	63.88	61.37	63.25	62.63	62.00	63.25	62.50
	SFDA-IT (Hou & Zheng, 2021a)	63.19	61.91	64.47	63.83	65.10	62.55	64.47	63.83	63.19	64.47	63.70
	SFDA-PFT (ours)	**69.36**	**67.96**	**70.76**	**70.06**	**71.46**	**68.66**	**70.76**	**70.06**	**69.36**	**70.76**	**69.92**
Oracle	Supervised fine-tuning	96.72	94.76	98.67	97.70	99.65	95.74	98.67	97.70	96.72	98.67	97.50

Comparison with State-of-the-Art Methods

BAH dataset (10 target subjects, 214 source subjects). Bold numbers indicate the best F1 score.

Setting	Methods	Sub-1	Sub-2	Sub-3	Sub-4	Sub-5	Sub-6	Sub-7	Sub-8	Sub-9	Sub-10	Average
Source-only	Source model	11.20	17.84	12.60	18.50	14.10	16.92	10.30	13.40	16.00	15.31	14.62
SFDA (model-based)	SHOT (Liang et al., 2020)	40.53	47.91	42.14	46.20	39.81	48.52	41.02	45.70	44.23	45.13	44.10
	NRC (Yang et al., 2021)	48.72	42.30	46.00	44.10	41.81	47.58	43.71	44.65	47.93	44.12	45.00
	TPDS (Tang et al., 2024)	41.22	46.30	44.01	42.54	47.82	40.95	45.53	43.29	47.18	42.23	44.20
	DSFDA (Sharafi et al., 2025)	49.10	44.70	47.51	42.92	50.23	45.30	46.70	47.90	41.82	49.84	46.10
SFDA (data-based)	SFIT (Hou & Zheng, 2021b)	56.83	50.91	54.72	52.10	57.54	49.82	55.91	51.23	58.12	51.40	52.90
	SFDA-IT (Hou & Zheng, 2021a)	48.50	55.71	50.81	53.95	47.21	54.12	49.03	52.64	50.32	**56.00**	51.80
	SFDA-PFT (ours)	**61.52**	**55.10**	**60.42**	**53.81**	**59.73**	**56.05**	**61.91**	**54.25**	**62.84**	54.70	**57.40**
Oracle	Supervised fine-tuning	96.20	92.81	95.70	94.25	96.53	93.91	95.14	94.72	92.53	97.01	94.88

Aff-Wild2 dataset (10 target subjects, 282 source subjects). Bold numbers indicate the best F1 score.

Setting	Methods	Sub-1	Sub-2	Sub-3	Sub-4	Sub-5	Sub-6	Sub-7	Sub-8	Sub-9	Sub-10	Average
Source-only	Source model (no adaptation)	18.70	19.60	20.50	20.00	21.00	20.50	21.40	22.30	20.00	21.00	20.50
SFDA (model-based)	SHOT (Liang et al., 2020)	33.77	34.67	35.57	35.07	36.07	35.57	36.47	37.37	35.07	36.07	35.57
	NRC (Yang et al., 2021)	34.24	35.14	36.04	35.54	36.54	36.04	36.94	37.84	35.54	36.54	36.04
	TPDS (Tang et al., 2024)	36.69	37.59	38.49	37.99	38.99	38.49	39.39	40.29	37.99	38.99	38.49
	DSFDA (Sharafi et al., 2025)	37.26	38.16	39.06	38.56	39.56	39.06	39.96	40.86	38.56	39.56	39.06
SFDA (data-based)	SFIT (Hou & Zheng, 2021b)	48.43	49.33	50.23	49.73	50.73	50.23	51.13	52.03	49.73	50.73	50.23
	SFDA-IT (Hou & Zheng, 2021a)	49.30	50.20	51.10	50.60	51.60	51.10	52.00	52.90	50.60	51.60	51.10
	SFDA-PFT (ours)	**52.66**	**53.56**	**54.46**	**53.96**	**54.96**	**54.46**	**55.36**	**56.26**	**53.96**	**54.96**	**54.46**
Oracle	Supervised fine-tuning	91.93	92.83	93.73	93.23	94.23	93.73	94.63	95.53	93.23	94.23	93.73

Computational Complexity

Comparison of SFDA models on BioVid dataset in terms of accuracy, number of iterations, and convergence time per batch (B=64).

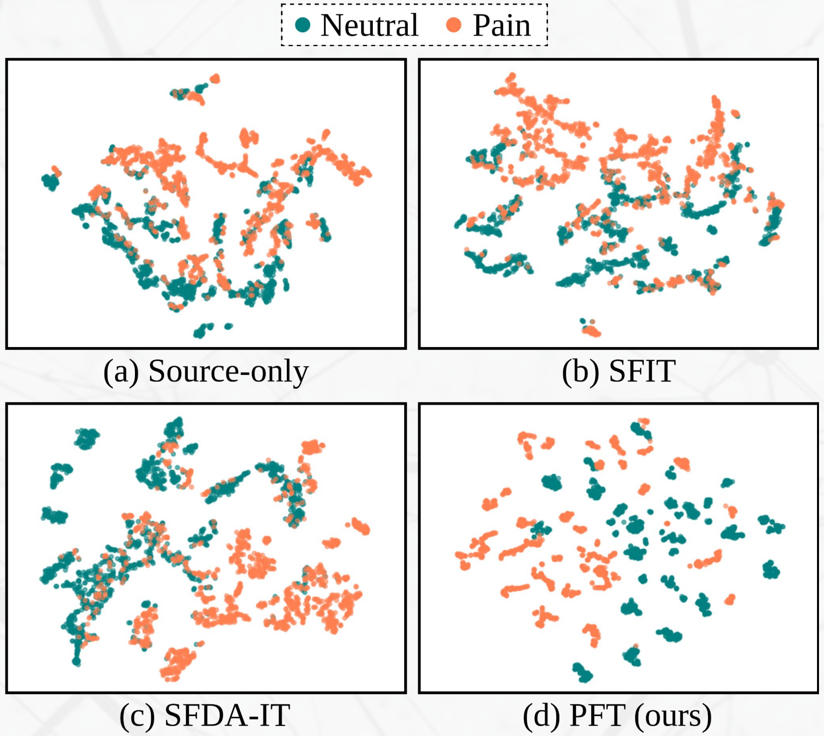
Method	ACC (%)	Iters	Time (s)
SFDA-DE (Liang et al., 2020)	62.88	1400	65.5
TPDS (Tang et al., 2024)	65.57	900	60.0
SHOT (Liang et al., 2020)	50.35	1155	54.0
NRC (Yang et al., 2021)	60.31	705	75.0
SFDA-PFT (ours)	**82.46**	**135**	**0.95**

Similarity of expression and identity branches on BioVid. Fixed expr. means the same expression, different subjects, while Fixed subj. means same subject, different expressions.

Branch	Fixed expr.	Fixed subj.
Expression	0.75	0.40
Identity	0.53	0.85

Visual Results

T-SNE of source vs. translated features for Sub-1 in BioVid dataset comparing feature-based (left) and image-based (right) translation.

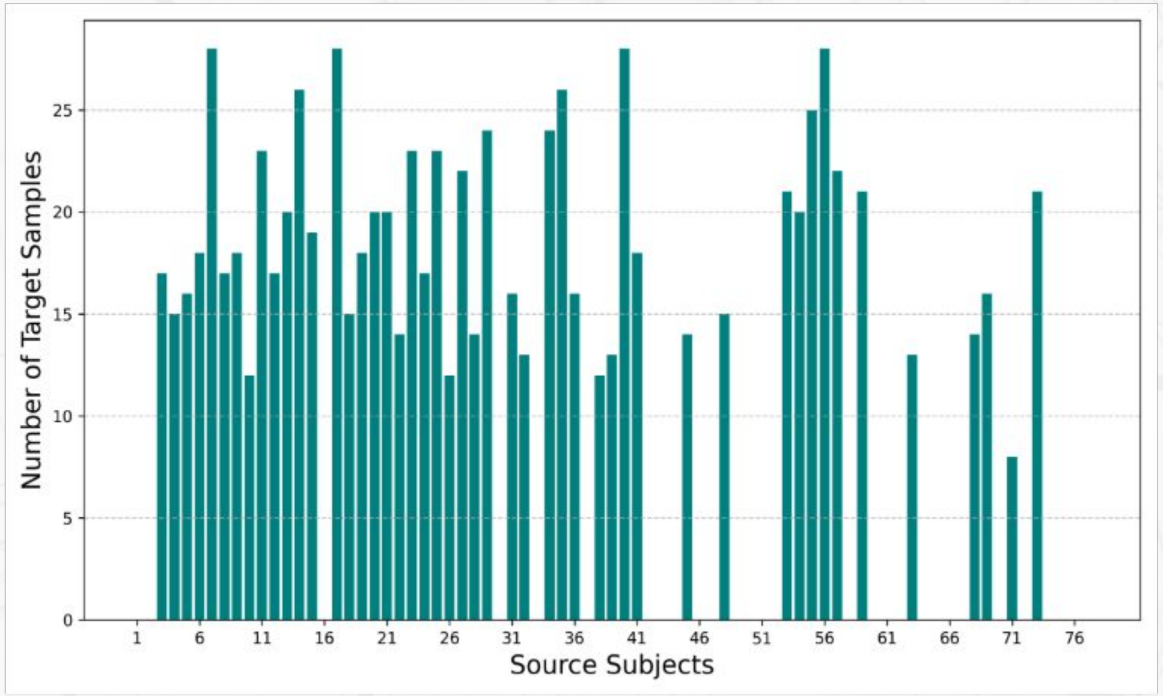


Source subject pairing on the BioVid dataset. (a) Examples of random, cosine-based, and landmark-based pairs. (b) Average accuracy, with landmark-based pairing performing best.

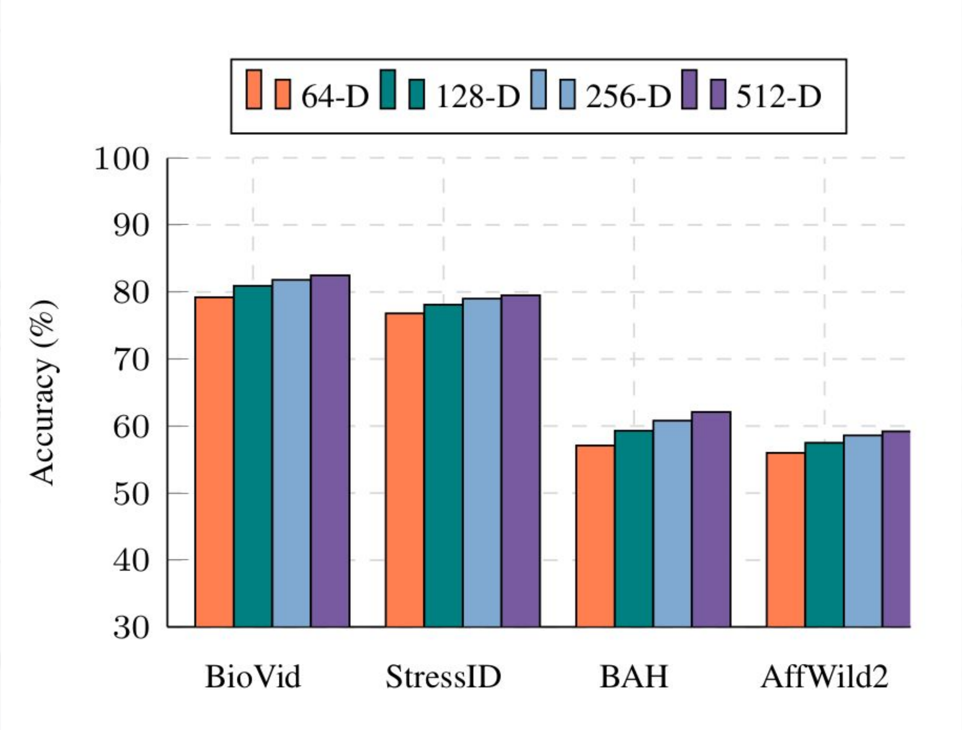
(a) **(b)**

Ablation Studies

Distribution of target samples for sub-1 in BioVid dataset across source subjects



SFDA-PFT accuracy across feature dimensions (64–512) on four datasets, showing performance gains with higher dimensions



Conclusion

- SFDA-PFT personalizes FER from only neutral target videos in feature space
- It avoids expressive target data and costly image generation
- Adaptation is lightweight, updating only a few translator layers
- Promising results on four video FER benchmarks with lower complexity

Code: https://github.com/MasoumehSharafi/SFDA-PFT

arXiv paper: https://arxiv.org/abs/2508.09202
"Personalized Feature Translation for Expression Recognition: An Efficient Source-Free Domain Adaptation Method"

For more discussions, please visit us at the **poster session**