AAAI-25 / IAAI-25 / EAAI-25

SLIP: Spoof-Aware One-Class Face Anti-Spoofing with Language Image Pretraining

Pei-Kai Huang[1], Jun-Xiong Chong[1], Cheng-Hsuan Chiang[1], Tzu-Hsien Chen[1], Tyng-Luh Liu[2], and Chiou-Ting Hsu[1]

[1]National Tsing Hua University, Taiwan [2]Academia Sinica, Taiwan

One-class face anti-spoofing (FAS)

❑ Face Anti-Spoofing (FAS)

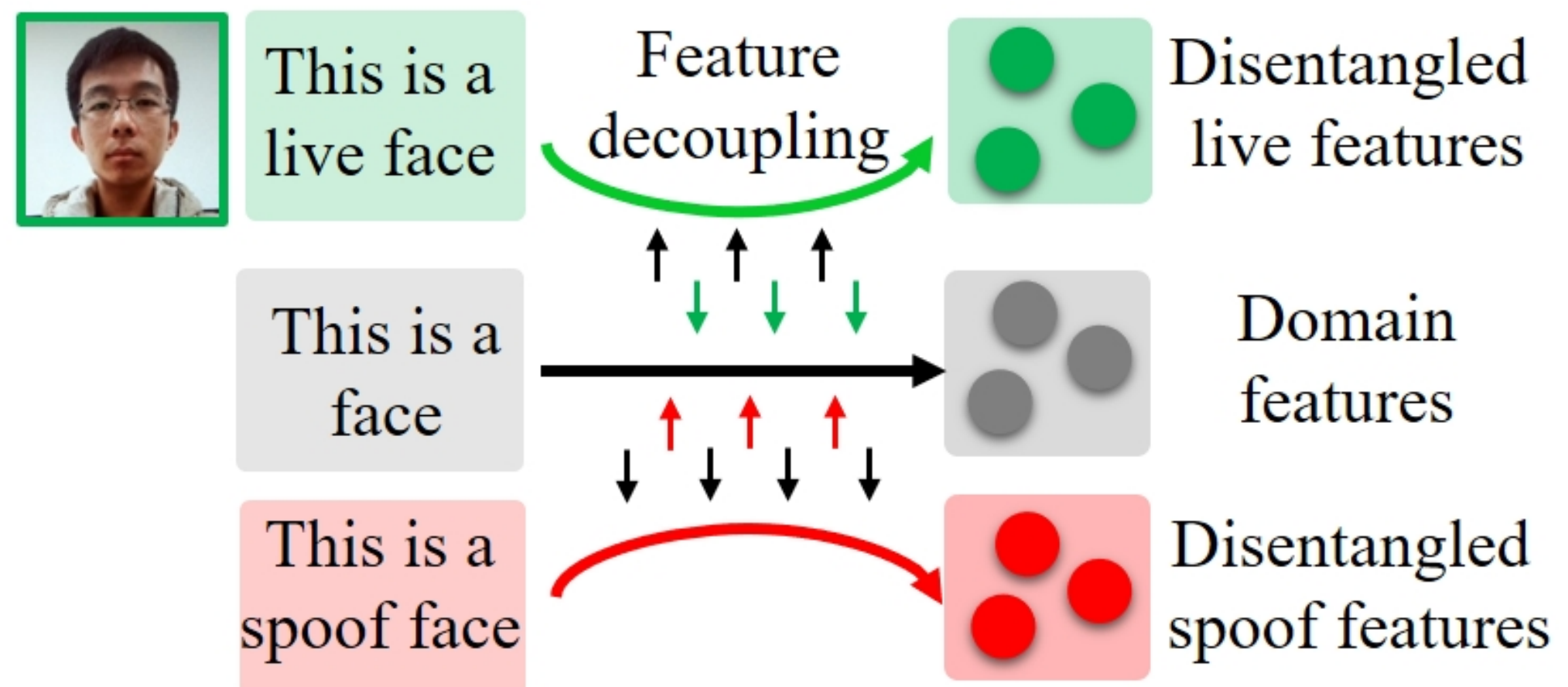
- To detect **facial spoof attacks**
  - Print attack, replay attack, 3D mask

❑ Challenges in one-class FAS

- Absence of training spoof images
- Similar visual characteristics between live and spoof faces
- Unseen spoof attacks
- Domain-entangled features

❑ Goals

- To learn **domain-disentangled** and **live/spoof discriminative features** in one-class FAS

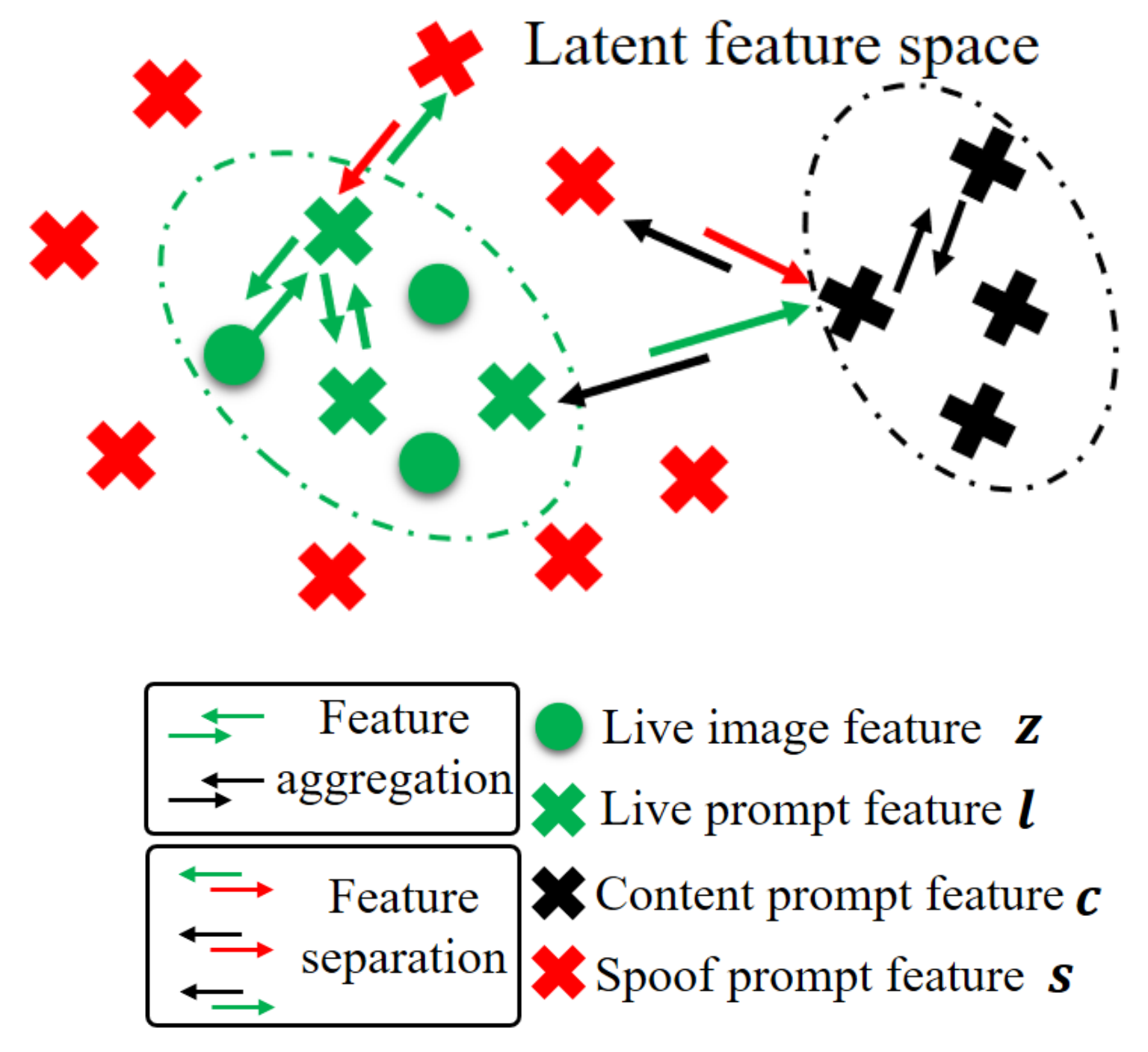


❑ Ideas

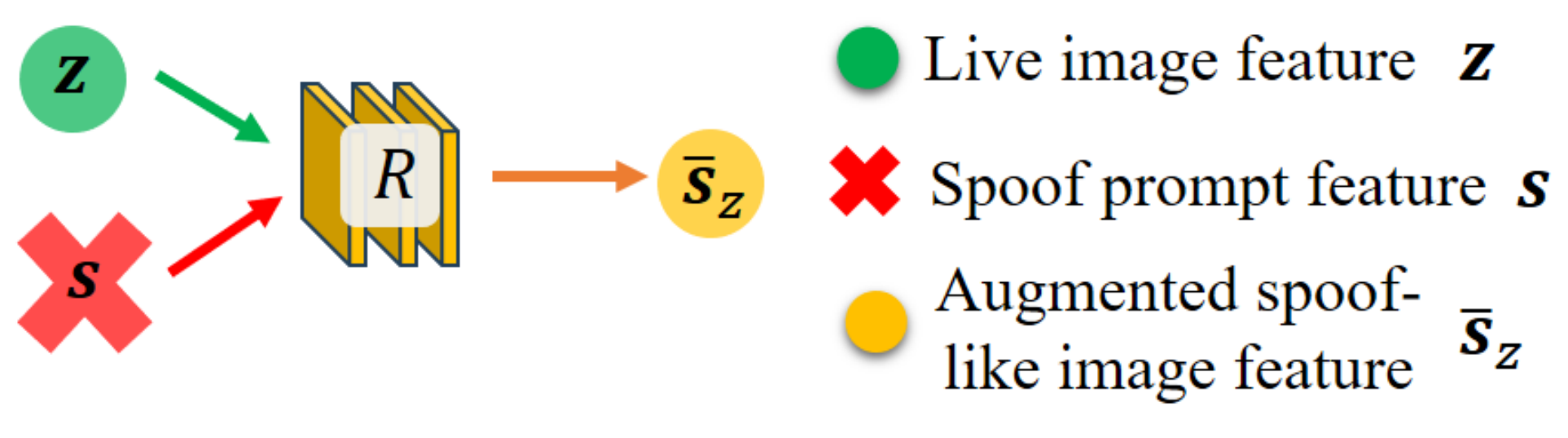
- Simulating spoof attacks via prompt learning

This is a spoof face modified by covering a live face with a photo.

- Disentangling domain information from live/spoof-discriminative features



- Augmenting spoof-like features



Spoof-aware one class face anti-spoofing with Language Image Pretraining

❑ Language-guided spoof cue map (SCM) estimation

- Zero SCM estimation from live Images $\mathbf{x}$ and live prompts $\mathbf{t}_l$

$$\mathcal{L}_L = \mathcal{L}_I + \mathcal{L}_T = \left(\sum\nolimits_{\mathbf{x}\in X}\|D(\mathbf{z}) - \mathbf{0}\|_2^2\right) + \left(\sum\nolimits_{\mathbf{t}_l\in T_l}\|D(\mathbf{l}) - \mathbf{0}\|_2^2\right)$$
$$= \left(\sum\nolimits_{\mathbf{x}\in X}\|D(E_I(\mathbf{x})) - \mathbf{0}\|_2^2\right) + \left(\sum\nolimits_{\mathbf{t}_l\in T_l}\|D(E_T(\mathbf{t}_l)) - \mathbf{0}\|_2^2\right)$$

- Nonzero SCM estimation from spoof prompts $\mathbf{t}_s$

$$\mathcal{L}_S = \sum\nolimits_{\mathbf{t}_s\in T_s,\tilde{\mathbf{m}}\in\mathcal{M}}\|D(\mathbf{s}) - \tilde{\mathbf{m}}\|_2^2 = \sum\nolimits_{\mathbf{t}_s\in T_s,\tilde{\mathbf{m}}\in\mathcal{M}}\|D(E_T(\mathbf{t}_s)) - \tilde{\mathbf{m}}\|_2^2$$

❑ Prompt-driven feature disentanglement

- Separation of live/spoof prompt features $\mathbf{l}$, $\boldsymbol{s}$ from content prompt features $\mathbf{c}$

$$\mathcal{L}_{FD} = -\sum_{i=1}^{N_c}\sum_{j,j\neq i}^{N_c}\left(\log\frac{\exp(\cos(\mathbf{c}_i,\mathbf{c}_j))}{\sum_{p=1}^{N_\mathbf{l}}\exp(\cos(\mathbf{c}_i,\mathbf{l}_p)) + \sum_{q=1}^{N_\mathbf{s}}\exp(\cos(\mathbf{c}_i,\mathbf{s}_q))}\right)$$
$$-\sum_{i=1}^{N_\mathbf{l}}\sum_{j\neq i}^{N_\mathbf{l}}\left(\log\frac{\exp(\cos(\mathbf{l}_i,\mathbf{l}_j))}{\sum_{k=1}^{N_\mathbf{s}}\exp(\cos(\mathbf{l}_i,\mathbf{s}_k))}\right)$$

- Alignment between live prompt features $\mathbf{l}$ and live image features $\boldsymbol{z}$

$$\mathcal{L}_{FA} = -\sum_{i=1}^{N_\mathbf{l}}\sum_{j=1}^{N_\mathbf{z}}\left(\log\frac{\exp(\cos(\mathbf{l}_i,\mathbf{z}_j))}{\sum_{k=1}^{N_\mathbf{s}}\exp(\cos(\mathbf{l}_i,\mathbf{s}_k))}\right)$$

❑ Spoof-like image feature augmentation

- Training the fusion model R to reconstruct the hybrid prompt features $\mathbf{h}$

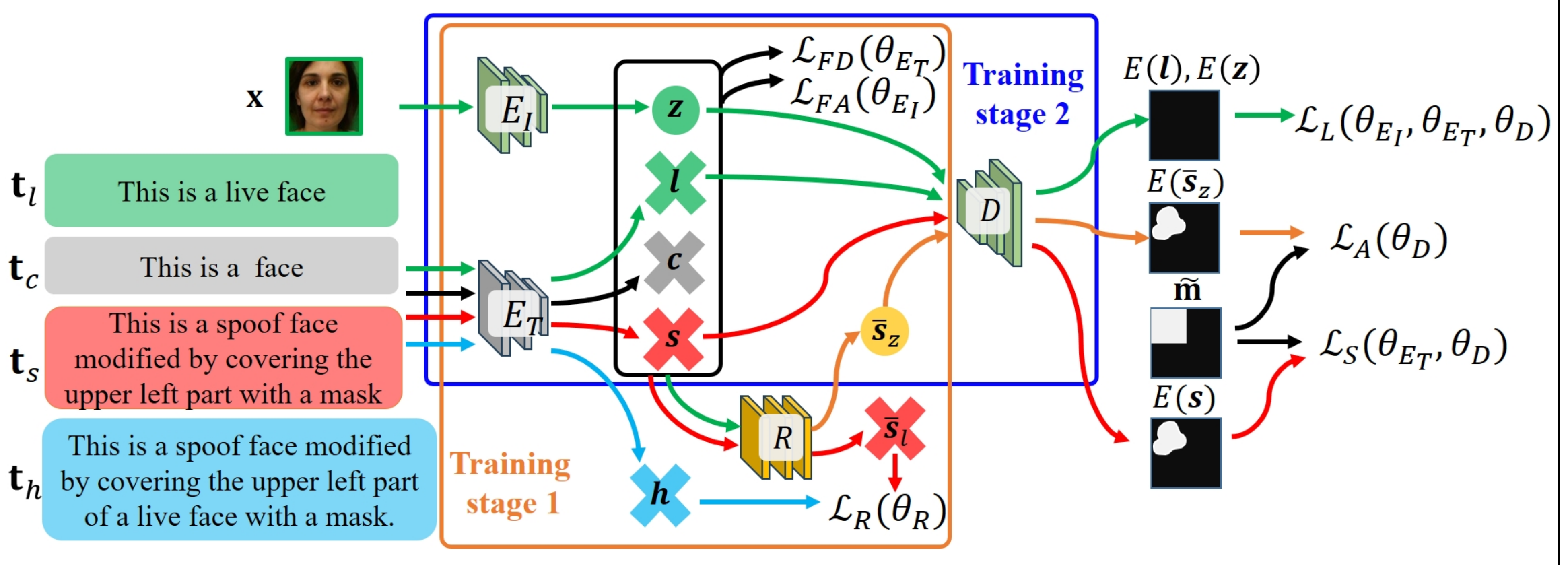
$$\mathcal{L}_R = \sum\nolimits_{\mathbf{t}_l\in T_l,\mathbf{t}_s\in T_s,\mathbf{t}_h\in T_h}\|R(\mathbf{l},\mathbf{s}) - \mathbf{h}\|_2^2 = \sum\nolimits_{\mathbf{t}_l\in T_l,\mathbf{t}_s\in T_s,\mathbf{t}_h\in T_h}\|\bar{\mathbf{s}}_l - \mathbf{h}\|_2^2$$

- Fusion of live image features $\boldsymbol{z}$ with spoof prompt features $\boldsymbol{s}$

$$\bar{\mathbf{s}}_z = R(\mathbf{z},\mathbf{s})$$

- Consistent SCM between the augmented features $\bar{\mathbf{s}}_z$ and the spoof prompt features $\mathbf{t}_s$

$$\mathcal{L}_A = \sum\nolimits_{\mathbf{x}\in X,\mathbf{t}_s\in T_s,\tilde{\mathbf{m}}\in\mathcal{M}}\|D(\bar{\mathbf{s}}_z) - \tilde{\mathbf{m}}\|_2^2$$


$\mathbf{x}$
$\mathbf{t}_l$
This is a live face
$\mathbf{t}_c$
This is a face
$\mathbf{t}_s$
This is a spoof face modified by covering the upper left part with a mask
$\mathbf{t}_h$
This is a spoof face modified by covering the upper left part of a live face with a mask.
E_I
E_T
z
l
c
s
h
R
$\bar{s}_z$
$\bar{s}_l$
D
$\mathcal{L}_{FD}(\theta_{E_T})$
$\mathcal{L}_{FA}(\theta_{E_I})$
$\mathcal{L}_R(\theta_R)$
Training stage 1
Training stage 2
$E(l), E(z)$
$E(\bar{s}_z)$
$\tilde{m}$
$E(s)$
$\mathcal{L}_L(\theta_{E_I},\theta_{E_T},\theta_D)$
$\mathcal{L}_A(\theta_D)$
$\mathcal{L}_S(\theta_{E_T},\theta_D)$

$$\boldsymbol{\theta}_{E_I}^*,\boldsymbol{\theta}_{E_T}^*,\boldsymbol{\theta}_D^* = \underset{\boldsymbol{\theta}_{E_I},\boldsymbol{\theta}_{E_T},\boldsymbol{\theta}_D}{\arg\min}\left(\mathcal{L}_L(\boldsymbol{\theta}_{E_I},\boldsymbol{\theta}_{E_T},\boldsymbol{\theta}_D) + \mathcal{L}_S(\boldsymbol{\theta}_{E_T},\boldsymbol{\theta}_D) + \lambda\mathcal{L}_{FD}(\boldsymbol{\theta}_{E_T}) + \lambda\mathcal{L}_{FA}(\boldsymbol{\theta}_{E_I}) + \mathcal{L}_A(\boldsymbol{\theta}_D)\right)$$

Experiments

Activation maps

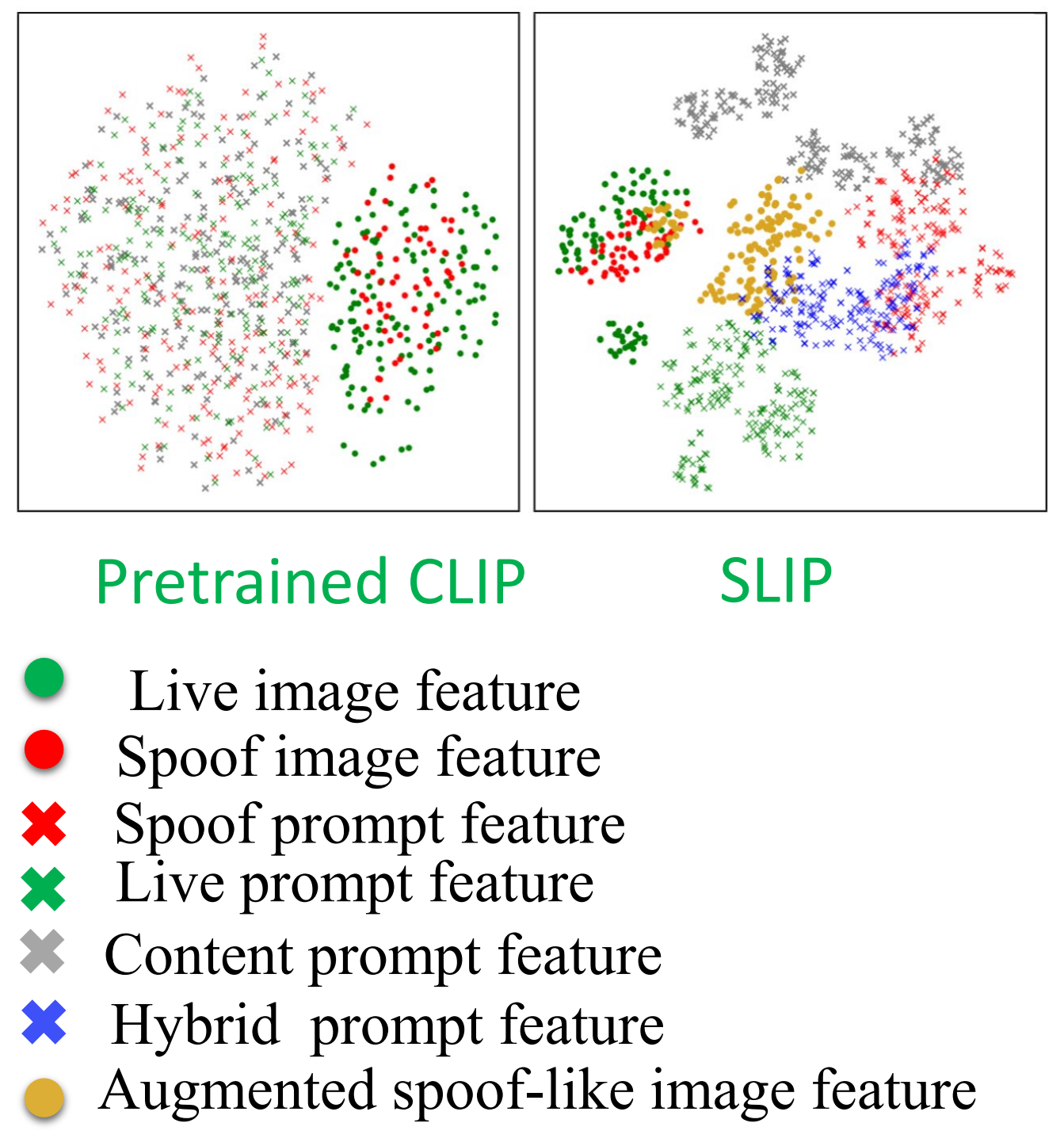
❑ Datasets

- OULU-NPU (O), CASIA-MFSD (C), MSU-MFSD (M), Idiap Replay-Attack (I), 3DMAD (D), HKBU-MARs (H), CASIA-SURF (U), and PADISI-Face (P)

❑ Evaluation Metrics

- APCER, BPCER, ACER, and HTER ↓
- AUC ↑

t-SNE visualization



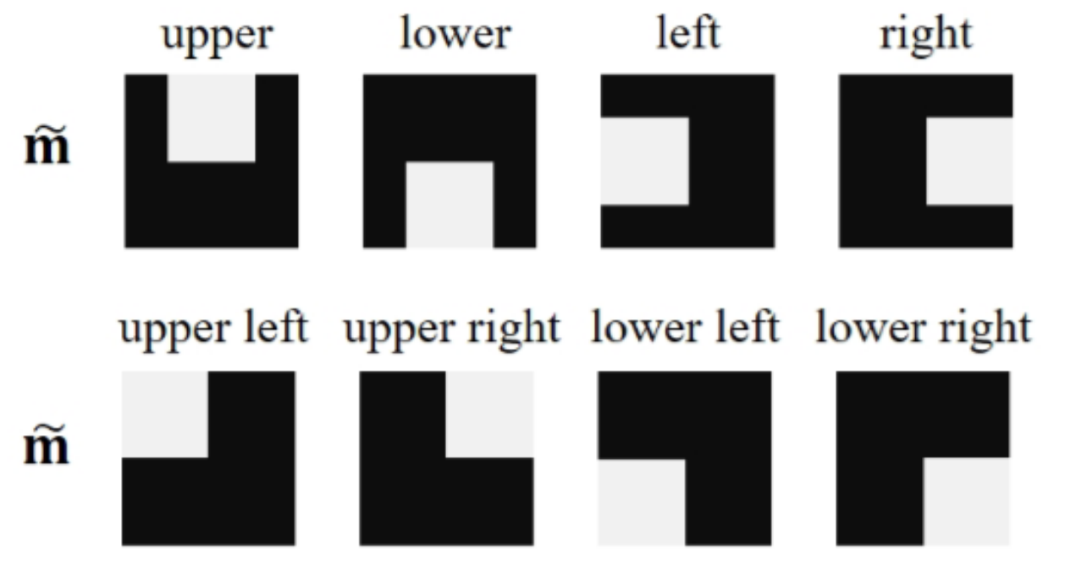
Intra-domain testing on Oulu

Type	Method	P.	APCER	BPCER	ACER
1-class	IQM-GMM (*ICB 18*)	1	75.35	18.56	46.95
	OC-fPAD (*IJCB 20*)		38.63	21.85	30.24
	OC-LCFAS (*Access 20*)		43.54	36.5	40.02
	AAE (*CCBR 21*)		47.13	26.67	36.9
	OC-SCMNet (*CVPR 24*)		20.83	26.15	23.49
	SLIP (Ours)		12.36	16.8	**14.58**
	IQM-GMM (*ICB 18*)	2	41.56	27.78	34.67
	OC-fPAD (*IJCB 20*)		51.81	19.83	35.82
	OC-LCFAS (*Access 20*)		72.19	18.5	45.35
	AAE (*CCBR 21*)		37.28	39.0	38.14
	OC-SCMNet (*CVPR 24*)		22.05	28.81	25.43
	SLIP (Ours)		22.16	23.18	**22.67**
	IQM-GMM (*ICB 18*)	3	57.17±16.79	16.5±6.95	36.83±5.35
	OC-fPAD (*IJCB 20*)		45.39±12.82	18.28±16.21	31.83±6.99
	OC-LCFAS (*Access 20*)		38.51±13.08	39.52±11.13	39.02±2.16
	AAE (*CCBR 21*)		26.62±13.67	52.93±16.09	39.77±3.74
	OC-SCMNet (*CVPR 24*)		27.10±12.57	20.55±11.12	23.83±3.14
	SLIP (Ours)		26.35±9.76	20.03±5.87	**23.19±2.86**
	IQM-GMM (*ICB 18*)	4	53.42±14.08	16.67±8.38	35.04±3.95
	OC-fPAD (*IJCB 20*)		60.25±16.49	10.67±10.37	35.46±5.43
	OC-LCFAS (*Access 20*)		36.91±10.24	20.5±8.01	28.07±5.32
	AAE (*CCBR 21*)		26.33±18.5	40.17±29.04	33.12±8.9
	OC-SCMNet (*CVPR 24*)		16.41±14.0	11.66±9.42	14.04±4.9
	SLIP (Ours)		15.02±3.84	10.9±4.66	**12.96±5.72**

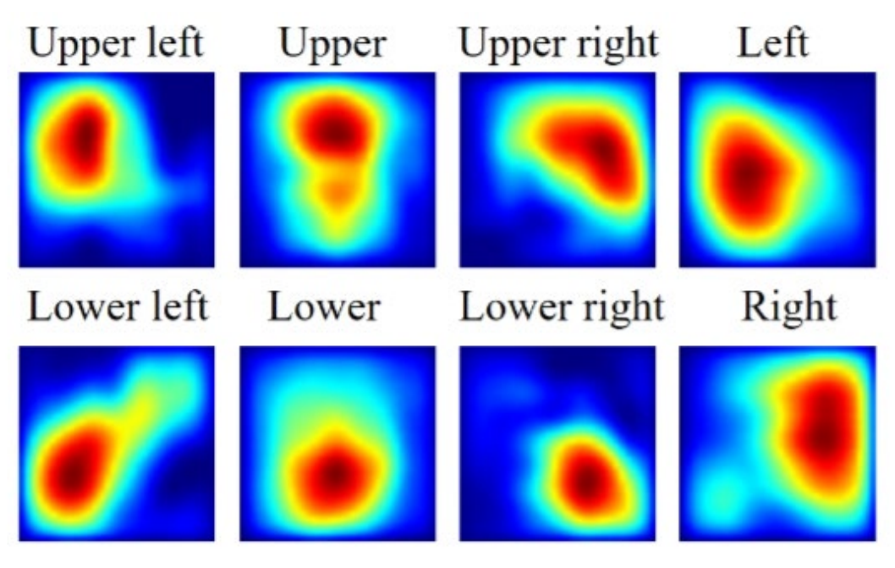
Unseen physical adversarial attack protocols

Type	Method	Funny eye		Paper grasses		Silicone 3Dmask	
		HTER	AUC	HTER	AUC	HTER	AUC
1-class	IQM-GMM(*ICB 18*)	30.11	68.82	15.01	88.82	22.53	80.33
	OC-fPAD(*IJCB 20*)	45.23	43.19	43.61	45.72	21.65	77.55
	OC-LCFAS(*Access 20*)	41.88	55.56	36.23	62.12	29.12	69.68
	AAE(*CCBR 21*)	45.92	46.25	31.20	72.03	27.00	67.26
	OC-SCMNet(*CVPR 24*)	28.99	60.46	14.33	92.26	8.40	84.27
	SLIP(Ours)	**19.77**	**81.56**	**7.02**	**97.61**	**3.86**	**98.32**

Pseudo spoof cue maps



Activation maps



Unseen attack protocols (print attacks + replay attacks)

Type	Method	OCI→M		OMI→C		OCM→I		ICM→O		#param.	FPS
		HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC		
1-class	IQM-GMM (*ICB 18*)	41.27	55.43	41.84	57.03	39.93	68.99	44.84	36.53	-	31
	OC-fPAD (*IJCB 20*)	39.51	60.65	32.64	74.86	38.25	73.01	39.62	69.71	145.03M	210
	OC-LCFAS (*Access 20*)	39.10	62.03	43.79	58.43	40.95	49.42	43.32	41.04	8.86M	388
	AAE (*CCBR 21*)	42.39	57.29	46.44	46.53	45.07	23.28	43.08	47.93	2.42M	816
	OC-SCMNet(*CVPR 24*)	24.05	75.53	28.02	76.92	21.36	87.29	34.37	69.87	5.92M	373
	SLIP (Ours)	**18.81**	**85.55**	**23.89**	**82.73**	**15.71**	**89.38**	**29.15**	**77.14**	171.72M	278

Unseen attack protocols (print attacks + replay attacks + 3D mask attacks)

Type	Method	Unseen 3D mask attacks				Unseen print attacks				Unseen replay attacks			
		OM → DHU		OCMI → DHU		OMD → OCMI		OCMIDHU → OCMI		OMD → OCMI		OCMIDHU → OCMI	
		HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
2-class	IADG (*CVPR 23*)	32.89	72.51	36.50	69.49	43.98	56.47	38.56	62.14	43.85	55.75	40.04	64.13
	SAFAS (*CVPR 23*)	38.22	63.75	34.48	65.33	**30.85**	**75.00**	40.09	63.16	39.12	64.99	38.45	66.69
1-class	IQM-GMM (*ICB 18*)	43.58	46.99	43.82	47.18	40.25	62.02	47.56	41.68	37.61	64.66	48.78	41.85
	OC-fPAD (*IJCB 20*)	39.35	61.86	42.19	57.47	41.59	61.56	40.41	63.83	48.06	42.45	46.87	41.26
	OC-LCFAS (*Access 20*)	41.74	56.43	41.64	55.11	46.17	53.45	48.29	50.30	41.32	59.08	46.45	53.71
	AAE (*CCBR 21*)	42.85	55.97	41.07	55.35	48.50	40.94	42.69	57.21	46.70	53.94	37.60	64.68
	OC-SCMNet (*CVPR 24*)	24.14	74.81	20.85	85.40	37.44	63.23	28.99	72.21	36.41	63.56	29.61	74.99
	SLIP (Ours)	**20.9**	**86.15**	**17.66**	**90.48**	31.29	74.85	**25.81**	**78.24**	**34.54**	**69.53**	**27.53**	**78.2**