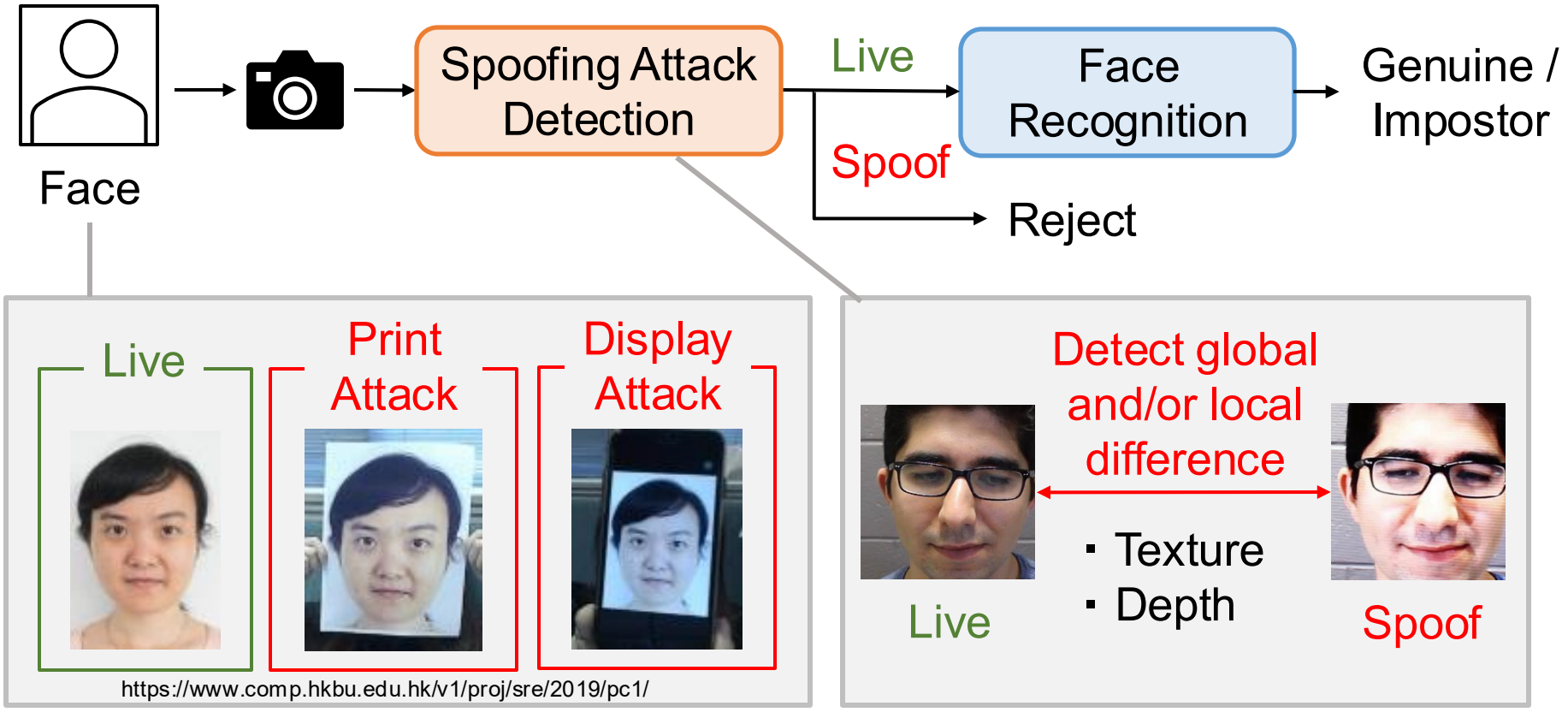
Leveraging Intermediate Features of Vision Transformer for Face Anti-Spoofing

Mika Feng[1], Koichi Ito[1], Takafumi Aoki[1]
Tetsushi Ohki[2], and Masakatsu Nishigaki[2]

1 Graduate School of Information Sciences, Tohoku University, Japan
2 Faculty of Informatics, Shizuoka University, Japan

Face anti-spoofing

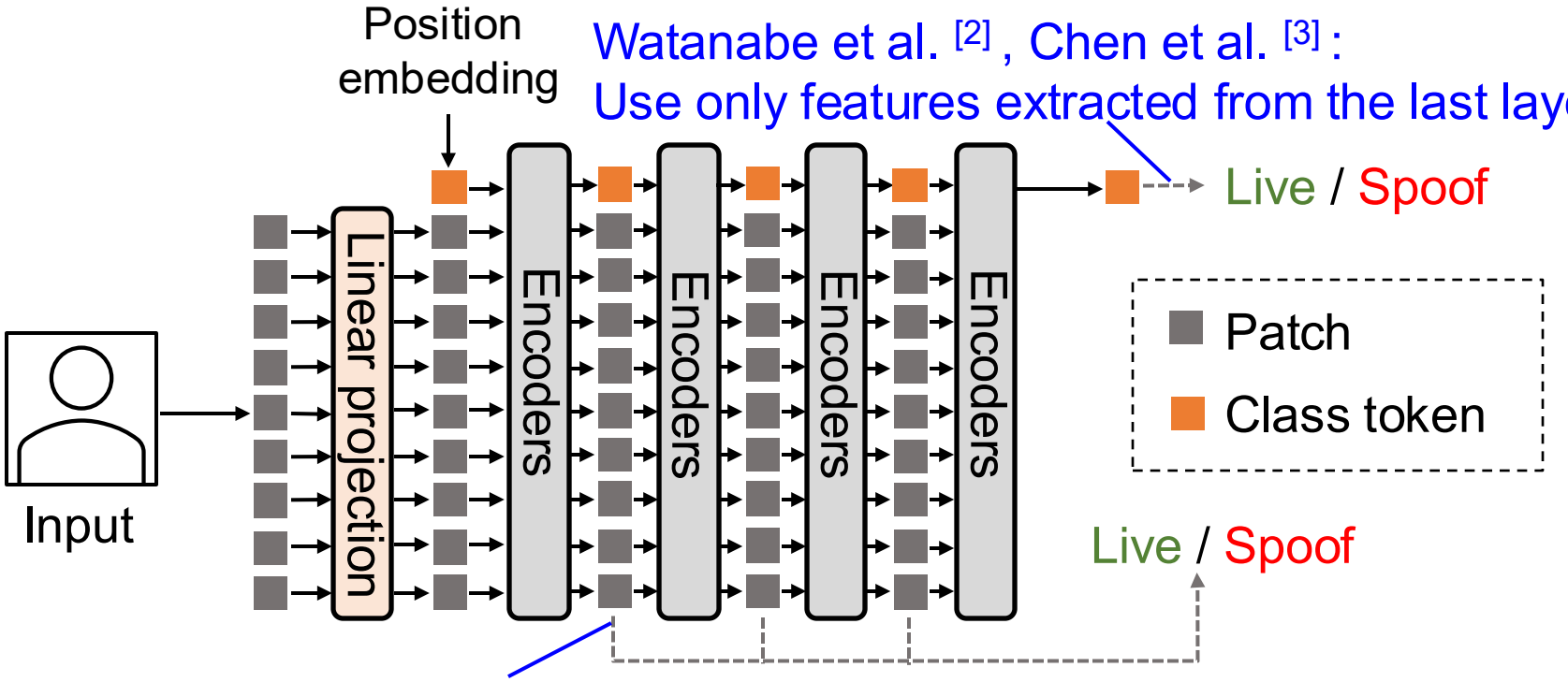
- Face recognition is robust against environmental changes
- If a face photo of a registered user is presented, a malicious person may bypass the authentication process illegally



Conventional methods using Vision Transformer (ViT)

- Local features can be extracted in the shallow layer and global features can be extracted in the deep layer

ViT [1]-based methods are not specialized for spoofing detection



[1] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” ICLR, 2021.

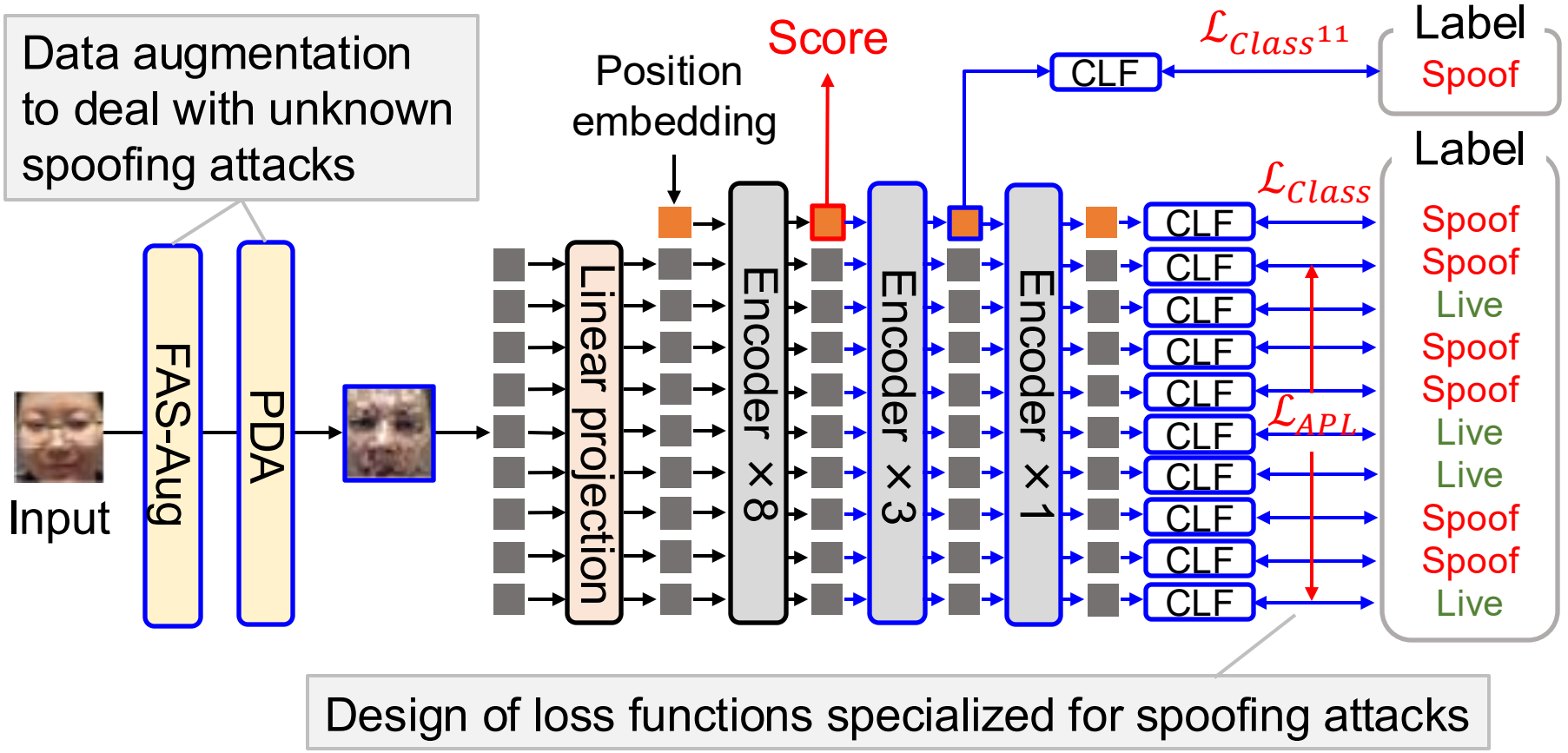
[2] K. Watanabe et al., “Spoofing attack detection in face recognition system using vision transformer with patch -wise data augmentation,” APSIPA, 2022.

[3] X. Chen et al., “Fine-grained annotation for face anti-spoofing,” arXiv, 2023.

[4] Z. Wang et al., “Face anti-spoofing using transformers with relation-aware mechanism,” IEEE T-BIOM, 2022.

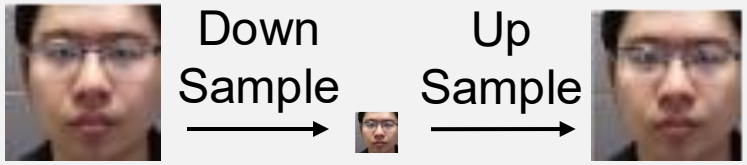
Proposed method

- Use intermediate features that balance local/global features and have high generalization performance without overfitting

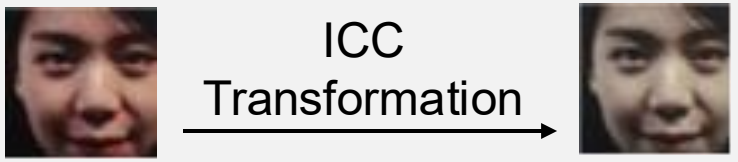


Face Anti-Spoofing data Augmentation (FAS-Aug)

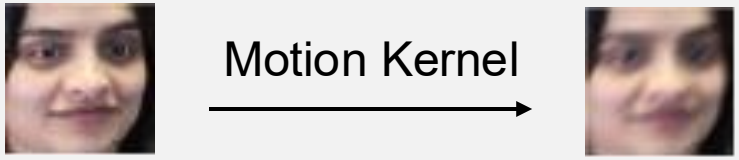
Low Resolution



Color Diversity



Hand Trembling



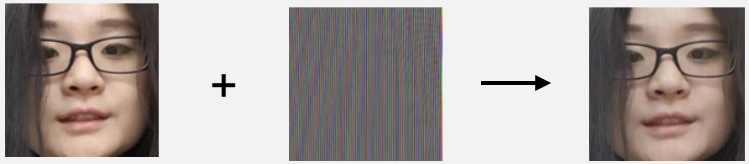
SFC Halftone

BN-Halftone

Color Distortion

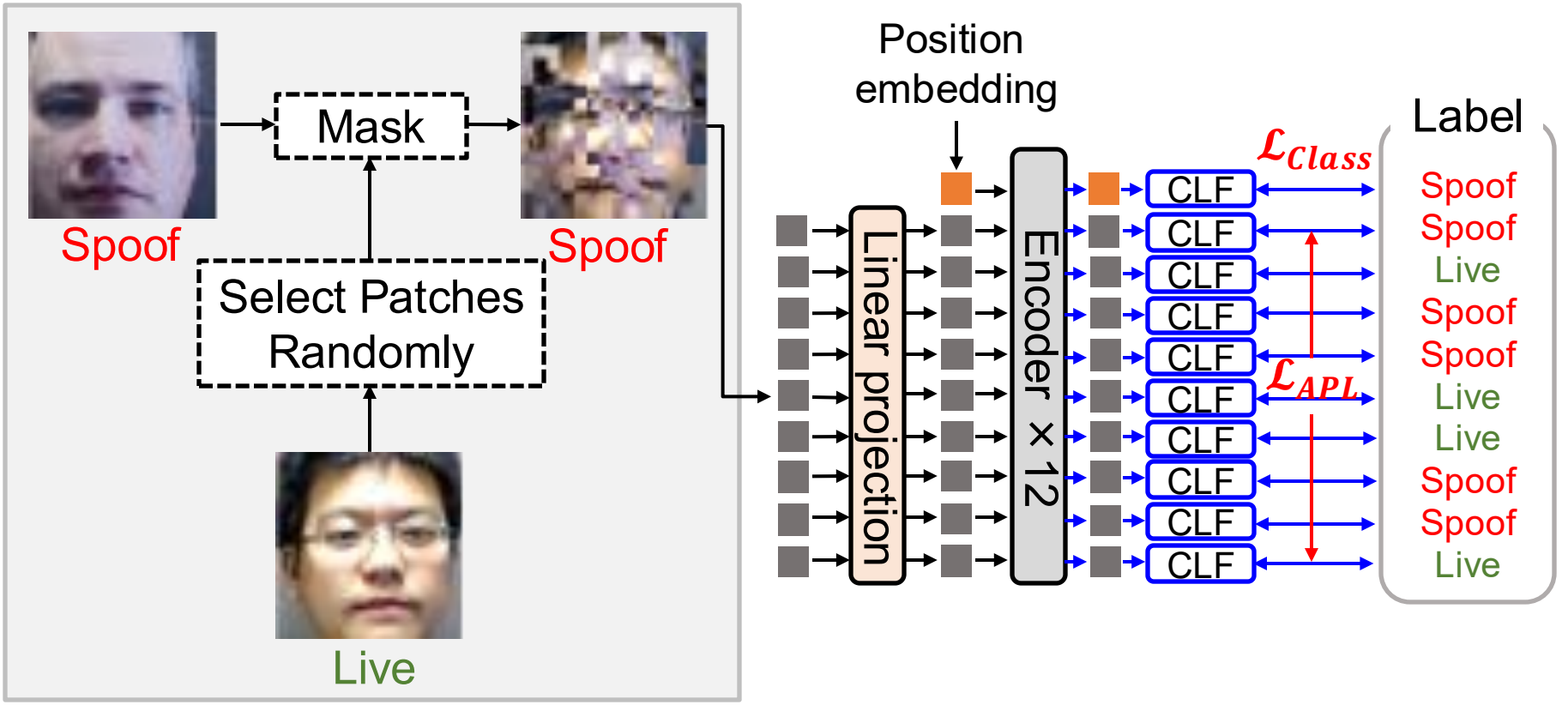
Specular Reflection

Moire Pattern



[5] R. Cai et al., “Towards Data-Centric Face Anti-spoofing: Improving Cross-Domain Generalization via Physics-Based Data Synthesis.” IJCV, 2024.

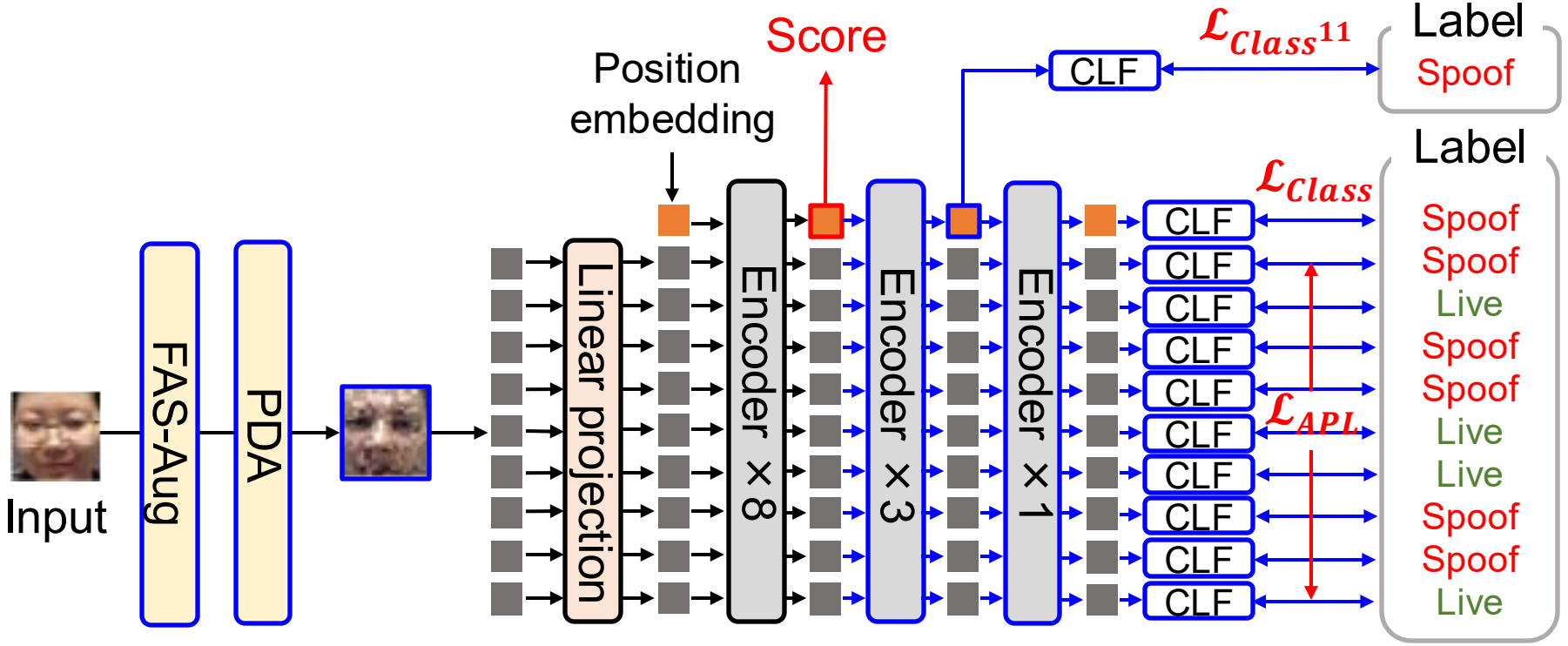
Patch-wise Data Augmentation (PDA)



- $\mathcal{L}_{APL}$ [2] : Take into account patch-wise spoof attack detection
- This makes detection more difficult, thereby enhancing the learning of the model

[2] K. Watanabe et al. “Spoofing attack detection in face recognition system using vision transformer with patch-wise data augmentation,” APSIPA, 2022.

Loss functions

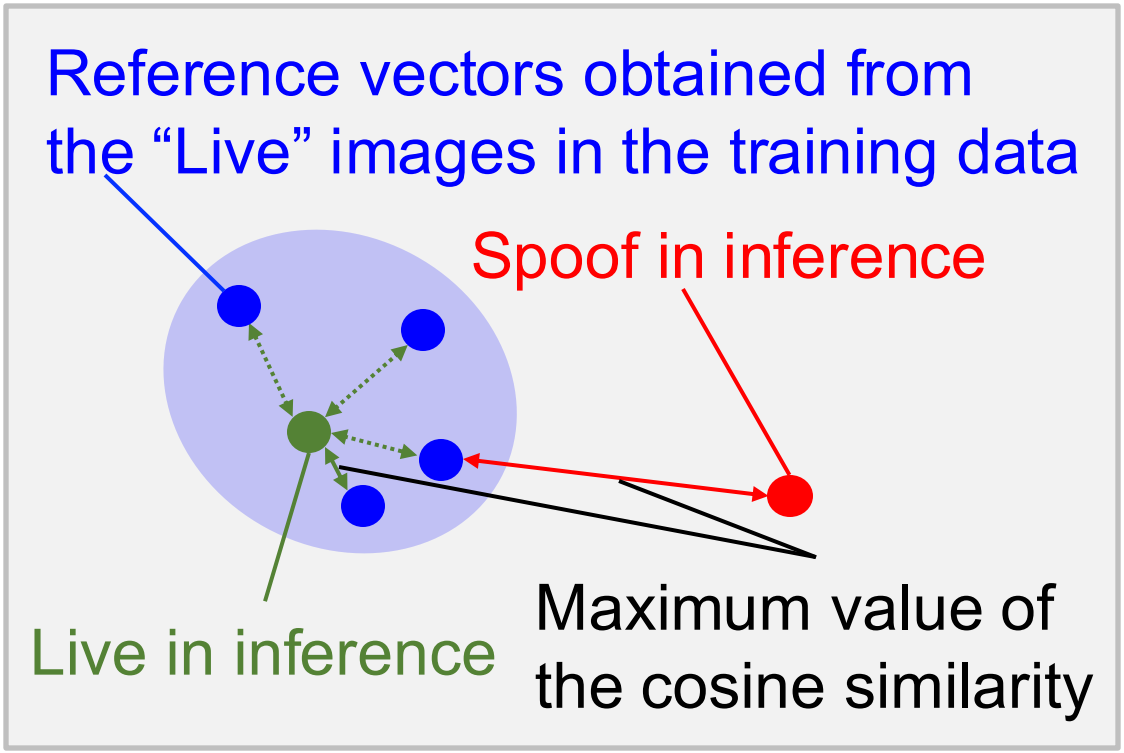


- $\mathcal{L}_{Class}{}^{11}$: Refine the features of the class token output from the 8th encoder block used to calculate the score
- Use L2-constrained softmax loss [6] to train the feature vectors equally without bias toward either “Live” or “Spoof”

[6] R. Ranjan et al., “L2-constrained softmax loss for discriminative face verification,” arXiv, 2017.

Score calculation

- Use the class token of the 8th encoder block to calculate score

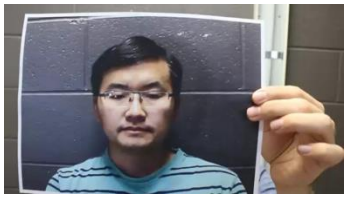


- If the score is greater than or equal to a threshold, the image is considered “Live,” otherwise, it is considered “Spoof.”

Dataset

- Use SiW [7] and OULU-NPU [8] datasets in the following experiments
- SiW [7] : Video of 165 subjects captured under varying lighting, head pose, and facial expression

Live



Print Attack

Display Attack

- Evaluation Protocol provided in SiW [7]

Prot.	Description
1	Changes in pose and facial expression
2	Types of display devices used in display attacks
3	Unknown spoofing attacks

[7] Y. Liu et al., “Learning deep models for face anti-spoofing: Binary or auxiliary supervision,” CVPR, 2018.
[8] Z. Boulkenafet et al., “OULU-NPU: A mobile face presentation attack database with real-world variations,” FG, 2017.

Experiments

i. Intermediate features of ViT [1] used for score calculation
   - The class token of the 8th encoder block balances local and global information without overfitting

ii. Intermediate features of ViT [1] used for loss calculation
   - The class token of the 8th encoder block is refined by adding constraints to that of the 11th encoder block

iii. Effectiveness of the loss functions: $\mathcal{L}_{APL}$ [2] and $\mathcal{L}_{Class}{}^{11}$
   - The combination of all loss functions are effective

iv. Effectiveness of the data augmentation methods: FAS-Aug [5] and PDA [2]
   - $P_{FAS-Aug} = 0.2$, $P_{PDA} = 0.2$

v. Comparison between the conventional and proposed methods using SiW [7] and OULU-NPU [8] dataset

v. Experimental results for SiW

Prot.	Method	APCER (%) ↓	BPCER (%) ↓	ACER (%) ↓
1	NAS-FAS [9]	0.07	0.17	0.12
	Watanabe [2]	0.11	0.08	0.10
	TransFAS [4]	0.00	0.00	0.00
	Proposed	0.1	0.08	0.09
2	NAS-FAS [9]	0.00±0.00	0.09±0.10	0.04±0.05
	Watanabe [2]	0.01±0.01	0.01±0.01	0.01±0.01
	TransFAS [4]	0.00±0.00	0.00±0.00	0.00±0.00
	Proposed	0.02±0.03	0.02±0.03	0.02±0.03
3	NAS-FAS [9]	1.58±0.23	1.46±0.08	1.52±0.13
	Watanabe [2]	3.07±2.75	3.07±2.75	3.07±2.75
	TransFAS [4]	1.95±0.40	1.92±0.11	1.94±0.26
	Proposed	0.83±0.13	0.83±0.14	0.83±0.13

[9] Z. Yu et al., "NAS-FAS: Static-dynamic central difference network search for face anti-spoofing," IEEE PAMI, 2020.