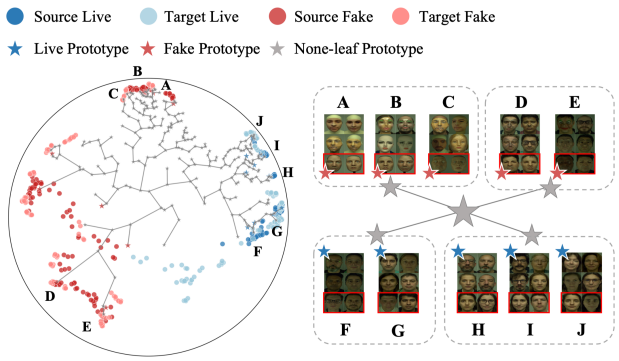
Rethinking Generalizable Face Anti-spoofing via Hierarchical Prototype-guided Distribution Refinement in Hyperbolic Space Supplementary Material

Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, Lizhuang Ma



(a) FlexibleMask (b) Examples of Prototype Samples

Figure S1. Visualization of the hierarchical structure in cross-attack setting when FlexibleMask is considered as the test dataset. We selected 10 prototypes with 4 samples in the source dataset and 2 samples from the target dataset (in the red box) as examples.

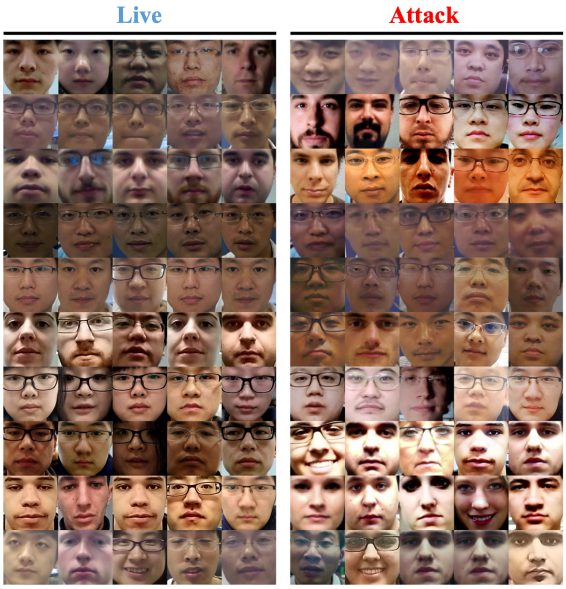


Figure S2. Visualization of all 20 prototypes with 5 samples in every row on O&C&M to I task.

1. Visualization of Cross-attack Setting

We display more visualization results on cross-attack settings to show the effectiveness of our HPL as shown in Figure S1. We observe that the live category exhibits a higher level of consistency in its general distribution compared to attack samples. This can be attributed to the fact that live samples are collected in well-conditioned environments. Additionally, we notice that the samples assigned to the same prototype within the live category tend to share similar environmental conditions. In the case of attack samples, we find that those with a lower LCA (Lowest Common Ancestor) tend to exhibit similarities in terms of the attack medium. For instance, Prototypes A, B, and C contain the samples significant differences compared to the live face samples, which utilize paper masks or rigid masks to fully conceal the human face. On the other hand, Prototypes D and E contain samples more similar to the live face samples, which only utilize makeup, glasses, and tattoos to conceal the partial face. In the target dataset, the samples with only the flexible mask are assigned to Prototypes A, B, and C since they exhibit significant differences compared to the live images, primarily due to the absence of a real human wearing. The samples of flexible masks worn by a human have been assigned to Prototypes D and E. This assignment is based on the fact that these samples contain live features such as hair, eyes, and ears, which closely resemble the characteristics observed in live images. It indicates that our HPL can generate a meaningful hierarchical structure and categorize the samples purposefully to align the features more accurately.

2. Visualization of all Prototypes

We further visualize the samples of all 20 prototypes on the O&C&M to I task in Figure S2. Each prototype displays 5 samples in the same row, which indicates that the samples assigned to the same prototype contain a high consistency. For live samples, all samples assigned to the same prototypes share a similar environment condition (*e.g.* lighting, contrast, white balance, *etc.*). All samples from the same row display a similar capture condition. For attack samples,

Backbone	Method	O&C&M to I		O&C&I to M		I&C&M to O		O&M&I to C	
		HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
ResNet-18	SSDG-R [6]	11.71	96.59	7.38	97.17	15.61	91.54	10.44	95.94
	SSAN-R [15]	8.88	96.79	6.67	98.75	13.72	93.63	10.00	**96.97**
	PatchNet [14]	13.40	95.67	7.10	98.46	11.82	95.70	11.33	94.58
	SA-FAS [13]	**6.58**	**97.54**	5.95	96.55	**10.00**	**96.23**	**8.78**	95.37
	Ours†	6.87	96.63	**2.50**	**98.82**	17.91	90.72	10.18	95.03
IADG	IADG [18]	10.62	94.50	5.41	98.19	8.86	97.14	8.70	96.44
	Ours	**7.63**	**96.12**	9.16	97.13	**7.84**	**97.32**	**4.63**	**99.21**

Table S1. Comparison with face anti-spoofing methods on four testing tasks for domain generalization on ResNet-18 backbone and the embedding part of IADG. †: Due to the lack of pretrain model with ImageNet on ResNet-18 in hyperbolic space, we train our method from scratch.

# None-leaf Prototype	O&M&I to C	
	HTER(%)	AUC(%)
32	15.92	92.02
64	14.44	93.00
128	12.40	93.15
256 (Ours)	**11.30**	**94.42**
512	13.33	93.02
1024	14.44	93.18

Table S2. Ablation study on the number of none-leaf prototypes.

the samples with similar attack types and attack mediums (*e.g.* print paper type, display devices, *etc.*) are easier to assign to the same prototype. This visualization indicates the effectiveness of our HPL module which forms meaningful prototypes with consistent sample features.

3. Comparison Result with Different Backbone

We have compared our proposed methods with the embedding part of DepthNet [10] and show promising results. To study our method with different backbones in this section, we compared our methods with different backbones like ResNet-18 [4] and embedding part of IADG [18]. The result is shown in Table S1. Due to the lack of the pretrain model on ImageNet in ResNet-18 backbone, our method trains with the ResNet-18 from scratch and still shows a comparable result. For the embedding part of the IADG, our method shows a promising result Compared to the IADG, which plugs the Categorical Style Assembly module to generate style-diversified samples for better generalization ability, our method shows a promising result, which indicates that the generated samples with diverse styles enrich the domain information and our HPDR can model this complex distribution in hyperbolic space with hierarchical structure and improve the generalization.

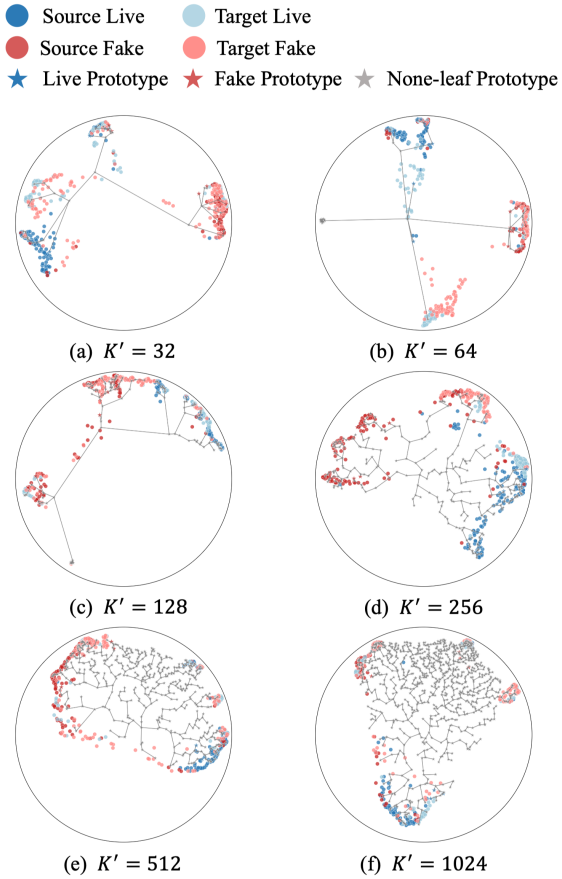


Figure S3. The feature distribution with different numbers of none-leaf prototypes in O&M&I to C task.

4. Comparison Result on WMCA Dataset

In this section, we apply the experiments on the recent dataset WMCA [2]. The experimental result of the grandtest (seen) and leave-one-out unseen attack protocols is shown in Table S3. We conduct our method with the

Method	Seen	Unseen							Average
		Glasses	Rigid Mask	Fake Head	Flexible Mask	Paper Mask	Print	Replay	
ResNet50 w/CCL [9]	30.69	16.80	17.62	24.67	4.76	9.51	19.03	15.37	15.39±6.51
Aux.(Depth) w/CCL [9]	30.62	**10.17**	27.32	40.00	7.41	11.67	16.11	12.76	17.92±11.66
CDCN w/CCL [9]	27.14	35.13	15.10	21.82	**7.18**	18.91	20.53	11.79	18.64±8.91
EPCR [16]	-	41.61	**2.78**	4.97	9.91	**2.47**	0.28	20.89	11.85±14.85
ViT [1, 7]	6.89	34.52	4.86	10.01	22.56	6.07	2.65	23.11	14.83±12.00
ViT-Hyp [3]	6.34	34.47	5.40	**4.00**	18.46	5.10	1.13	17.90	12.35±11.93
ViT-Hyp-HCL [3]	**5.62**	26.64	3.88	4.35	20.68	3.47	0.91	20.73	11.52±10.68
Ours	6.32	23.97	3.78	4.68	17.90	11.72	**0.00**	**10.89**	**10.42±8.44**

Table S3. Unimodal results of seen and unseen protocols on WMCA dataset. The values ACER(%) reported on testing sets are obtained with thresholds computed for BPCER=1% on development sets. The best results are bolded.

	Method	CS → W			SW → C			CW → S			Avg.
		HTER	AUC	TPR@ FPR=1%	HTER	AUC	TPR@ FPR=1%	HTER	AUC	TPR@ FPR=1%	HTER
0-shot	ViT [5]	7.98	97.97	73.61	11.13	95.46	47.59	13.35	94.13	49.97	10.82
	FLIP-V [12]	6.13	97.84	50.26	10.89	95.82	53.93	12.48	94.43	53.00	9.83
	FLIP-IT [12]	4.89	98.65	59.14	10.04	96.48	59.4	15.68	91.83	43.27	10.2
	FLIP-MCL [12]	4.46	99.16	83.86	9.66	96.69	59.00	11.71	95.21	57.98	8.61
	Ours	**4.39**	**99.26**	**87.78**	**2.11**	**99.82**	**95.13**	14.57	92.76	50.93	**7.02**

Table S4. Evaluation of cross-domain on WCS benchmark, between CASIA-SURF (**S**), CASIA-CeFA (**C**), and WMCA (**W**). We report the mean HTER, AUC, and TPR@FPR=1%

ResNet18 backbone and obtain a promising result on both seen and unseen settings even compared to the methods with the ViT backbone, which indicates the effectiveness of our method.

5. Comparison Result on CSW Benchmark

We also conduct the experiments on a more complex cross-domain setting, CSW benchmark, as shown in Table S4. We utilize three public datasets including CASIA-SURF (denoted as **S**) [17], CASIA-CeFA (denoted as **C**) [8], and WMCA (denoted as **W**) [2]. We train the model with two of the dataset and test on the rest one, with three settings including CS→W, SW→C, and CW→S. We follow the previous setting in FLIP [12] and apply our method with the same backbone of FLIP as the embedding part of the CLIP [11]. Our method shows a good result even in the complex cross-domain setting, which indicates that our hierarchical learning improves the generalization ability.

6. Ablation Study on None-leaf Prototypes

In this section, we explore the influence of none-leaf prototypes. We conduct the experiment on O&M&I to C task and show the result in Table S2. We find that, when the prototypes increase first, the performance is better, the reason is that, more none-leaf prototypes facilitate the construction of the hierarchical structure. When the none-leaf prototype is still increasing, the performance degrades a little. It shows

$\mathcal{L}_{PP-LCA}$	$\mathcal{L}_{P-Origin}$	$\mathcal{L}_{PP-leaf}$	O&M&I to C	
			HTER	**AUC**
			14.07	92.04
✓			12.40	93.14
	✓		13.33	92.21
		✓	13.98	92.63
✓	✓		12.01	94.01
✓		✓	12.00	93.90
	✓	✓	13.14	92.72
✓	✓	✓	**11.30**	**94.43**

Table S5. Ablation study on loss function of $\mathcal{L}_{PP}$.

that too many none-leaf prototypes may make it harder to optimize an effective solution. Also, more none-leaf prototypes will introduce more parameters which is computation-cost for training. We visualize the feature distributions with different none-leaf prototypes in Figure S3. When none-leaf prototype number is small, HPL faces difficulty in forming an elaborate hierarchical structure. With the increase of the none-leaf prototype number, the hierarchical structure is more clear. Nevertheless, with too many none-leaf prototypes, HPL cannot form an effective hierarchical structure due to the difficulty in optimizing too many prototypes. Also, most of the prototypes are underutilized as shown in Figure S3 (e-f), which may hinder the performance.

7. Ablation Study on Prototype-prototype Loss

Further, we apply the ablation on $\mathcal{L}_{PP-LCA}$ (Eq.15), $\mathcal{L}_{P-Origin}$ (Eq.16) and $\mathcal{L}_{PP-leaf}$ (Eq.17) three components of $\mathcal{L}_{PP}$ in Table S5. We find that: 1) $\mathcal{L}_{PP-LCA}$ plays the most important role with a large gain in performance, which indicates the importance of building a hierarchical structure. 2) With $\mathcal{L}_{PP-leaf}$ and $\mathcal{L}_{P-Origin}$, the leaf prototypes are more representative and perform better, which shows these loss functions improve the distribution of the samples. 3) With only $\mathcal{L}_{PP-leaf}$ or $\mathcal{L}_{P-Origin}$, for lack of hierarchical structure, the performance improves little. 4) Our $\mathcal{L}_{PP}$ with all three components performs best, indicating all losses' effectiveness.

References

[1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. 3

[2] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE transactions on information forensics and security*, 15:42–55, 2019. 2, 3

[3] Shuangpeng Han, Rizhao Cai, Yawen Cui, Zitong Yu, Yongjian Hu, and Alex Kot. Hyperbolic face anti-spoofing. *arXiv preprint arXiv:2308.09107*, 2023. 3

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2

[5] Hsin-Ping Huang, Deqing Sun, Yaojie Liu, Wen-Sheng Chu, Taihong Xiao, Jinwei Yuan, Hartwig Adam, and Ming-Hsuan Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*, pages 37–54. Springer, 2022. 3

[6] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *CVPR*, 2020. 2

[7] Chen-Hao Liao, Wen-Cheng Chen, Hsuan-Tung Liu, Yi-Ren Yeh, Min-Chun Hu, and Chu-Song Chen. Domain invariant vision transformer learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6098–6107, 2023. 3

[8] Ajian Liu, Zichang Tan, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1179–1187, 2021. 3

[9] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing, Yanyan Liang, et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE TIFS*, 2022. 3

[10] Yaojie Liu*, Amin Jourabloo*, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *CVPR*, Salt Lake City, UT, 2018. 2

[11] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021. 3

[12] Koushik Srivatsan, Muzammal Naseer, and Karthik Nandakumar. Flip: Cross-domain face anti-spoofing with language guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19685–19696, 2023. 3

[13] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *CVPR*, pages 24563–24574, 2023. 2

[14] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. In *CVPR*, pages 20281–20290, 2022. 2

[15] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *CVPR*, pages 4123–4133, 2022. 2

[16] Zezheng Wang, Zitong Yu, Xun Wang, Yunxiao Qin, Jiahong Li, Chenxu Zhao, Zhen Lei, Xin Liu, Size Li, and Zhongyuan Wang. Consistency regularization for deep face anti-spoofing. *IEEE TIFS*, 2023. 3

[17] Shifeng Zhang, Ajian Liu, Jun Wan, Yanyan Liang, Guodong Guo, Sergio Escalera, Hugo Jair Escalante, and Stan Z Li. Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2):182–193, 2020. 3

[18] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Ran Yi, Shouhong Ding, and Lizhuang Ma. Instance-aware domain generalization for face anti-spoofing. In *CVPR*, pages 20453–20463, 2023. 2