# Supplementary Material
# FLIP: Cross-domain Face Anti-spoofing with Language Guidance

Koushik Srivatsan Muzammal Naseer Karthik Nandakumar
Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI)
Abu Dhabi, United Arab Emirates
{koushik.srivatsan, muzammal.naseer, karthik.nandakumar}@mbzuai.ac.ae

In this supplementary material, we provide the experimental results for extending the FLIP framework to the 5-shot setting in section 1. In section 2, we provide the results of the statistical significance test. In section 3, we compare the computational complexity and network parameters between all the methods in the FLIP framework. In section 4, we present the results of evaluating on unseen spoof type.

## 1. Performance of FLIP in 5-shot setting

Following [4], we evaluate the FLIP framework under the 5-shot setting, where 5 labeled samples from the target domain are available during training to help bridge the domain gap. Tables 1, 2, and 3 report the cross-domain 0-shot and 5-shot performance of Protocol 1, 2, and 3, respectively.

**5-shot performance in Protocol 1:** We observe that all the 3 methods from the FLIP framework outperform the baseline 5-shot performance. Notably, for the O protocol (where the target samples have higher image resolution and are 4 times larger than all the source domains combined), we observe a large HTER gain of +3.85%. This demonstrates that our method is able to effectively adapt to larger unknown domains with very few samples ($\approx$ 0.16% of target domain samples in O).

**5-shot performance in Protocol 2:** Similar to **Protocol 1**, we observe that our framework outperforms the baseline 5-shot methods by a huge margin of +2.43% in terms of average HTER. Notably, for the C and S protocols (which contain more than 1000 identities and have large illumination variations), we observe HTER gains of +3.87% and +5.4% respectively. This demonstrates the effectiveness of our method in adopting to unknown distributions containing diverse samples, with just a few labeled samples (0.08% for C and 0.1% for S).

**5-shot performance in Protocol 3:** To make a fair comparison, we implement the baseline ViTAF* method [4] and extend it to Protocol 3 under the 5-shot setting. We observe that the performance of the FLIP framework in the 5-shot setting outperforms its 0-shot counterpart. Additionally, the 5-shot FLIP framework also outperforms 5-shot ViTAF* by a margin of +2.26% (HTER). This corroborates our previous observations on our approach's effectiveness in adapting to unknown domains with a few labeled samples.

## 2. Statistical Significance Test

Most prior works in cross-domain FAS simply report the best result over a single trial. However, a fair comparison of different methods is possible only when the statistical variations are taken into account. Hence, we run each of our experiments 5 times with different random seeds and report the mean and standard deviation of all the metrics in Tables 1, 2, and 3. For each of the three protocols, we observe that the standard deviation of the proposed method is low, indicating stable performance across multiple runs.

Furthermore, for **Protocol 1** and **Protocol 2**, we perform a one-sided pair-wise t-test to evaluate whether the proposed method outperforms the baseline. Specifically, we compare the proposed FLIP-MCL against ViT in the 0-shot setting and against ViTAF* [4] in the 5-shot setting. The null hypothesis is that there is no statistically significant difference between FLIP-MCL and the baseline, while the alternate hypothesis is that FLIP-MCL is better. In **Protocol 1**, we find that the null hypothesis is rejected in three out of four scenarios, failing only for M (for both the 0-shot and 5-shot setting). For **Protocol 2**, the null hypothesis is rejected for all three scenarios in the 0-shot setting. However, for the 5-shot setting, the null hypothesis is rejected for two out of three scenarios failing only in W. These results clearly demonstrate that FLIP-MCL is superior to the baseline methods and the better generalization performance is not due to cherry picking of best trials.

Table 1. Extending evaluation of cross-domain performance in Protocol 1 from 0-shot to 5-shot. We evaluate between MSU-MFSD (**M**), CASIA-MFSD (**C**), Replay Attack (**I**), and OULU-NPU (O). We run each experiment 5 times under different seeds and report the mean HTER, AUC, and TPR@FPR=1%, along with their standard deviation (shown in brackets under the mean scores).

| | **Method** | **OCI → M** | | | **OMI → C** | | | **OCM → I** | | | **ICM → O** | | | **Avg.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HTER | AUC | TPR@ FPR=1% | HTER | AUC | TPR@ FPR=1% | HTER | AUC | TPR@ FPR=1% | HTER | AUC | TPR@ FPR=1% | HTER |
| 0-shot | MADDG (CVPR' 19) [12] | 17.69 | 88.06 | – | 24.50 | 84.51 | – | 22.19 | 84.99 | – | 27.98 | 80.02 | – | 23.09 |
| | MDDR (CVPR' 20) [17] | 17.02 | 90.10 | – | 19.68 | 87.43 | – | 20.87 | 86.72 | – | 25.02 | 81.47 | – | 20.64 |
| | NAS-FAS (TPAMI' 20) [21] | 16.85 | 90.42 | – | 15.21 | 92.64 | – | 11.63 | 96.98 | – | 13.16 | 94.18 | – | 14.21 |
| | RFMeta (AAAI' 20) [13] | 13.89 | 93.98 | – | 20.27 | 88.16 | – | 17.30 | 90.48 | – | 16.45 | 91.16 | – | 16.97 |
| | $D^2$AM (AAAI' 21) [1] | 12.70 | 95.66 | – | 20.98 | 85.58 | – | 15.43 | 91.22 | – | 15.27 | 90.87 | – | 16.09 |
| | DRDG (IJCAI' 21) [11] | 12.43 | 95.81 | – | 19.05 | 88.79 | – | 15.56 | 91.79 | – | 15.63 | 91.75 | – | 15.66 |
| | Self-DA (AAAI' 21) [19] | 15.40 | 91.80 | – | 24.50 | 84.40 | – | 15.60 | 90.10 | – | 23.10 | 84.30 | – | 19.65 |
| | ANRL (ACM MM' 21) [10] | 10.83 | 96.75 | – | 17.85 | 89.26 | – | 16.03 | 91.04 | – | 15.67 | 91.90 | – | 15.09 |
| | FGHV (AAAI' 21) [9] | 9.17 | 96.92 | – | 12.47 | 93.47 | – | 16.29 | 90.11 | – | 13.58 | 93.55 | – | 12.87 |
| | SSDG-R (CVPR' 20) [5] | 7.38 | 97.17 | – | 10.44 | 95.94 | – | 11.71 | 96.59 | – | 15.61 | 91.54 | – | 11.28 |
| | SSAN-R (CVPR' 22) [20] | 6.67 | 98.75 | – | 10.00 | 96.67 | – | 8.88 | 96.79 | – | 13.72 | 93.63 | – | 9.80 |
| | PatchNet (CVPR' 22) [15] | 7.10 | 98.46 | – | 11.33 | 94.58 | – | 13.40 | 95.67 | – | 11.82 | 95.07 | – | 10.90 |
| | GDA (ECCV' 22) [23] | 9.20 | 98.00 | – | 12.20 | 93.00 | – | 10.00 | 96.00 | – | 14.40 | 92.60 | – | 11.45 |
| 0-shot | DiVT-M (WACV' 23) [8] | 2.86 | 99.14 | – | 8.67 | 96.62 | – | 3.71 | 99.29 | – | 13.06 | 94.04 | – | 7.07 |
| | ViT (ECCV' 22) [4] | 1.58 | 99.68 | 96.67 | 5.70 | 98.91 | 88.57 | 9.25 | 97.15 | 51.54 | 7.47 | 98.42 | 69.30 | 6.00 |
| 0-shot | FLIP-V | 3.79 *(1.40)* | 99.31 *(0.31)* | 87.99 *(6.09)* | 1.27 *(0.85)* | 99.75 *(0.18)* | 95.85 *(5.53)* | 4.71 *(2.39)* | 98.80 *(0.85)* | 75.84 *(16.53)* | 4.15 *(0.56)* | 98.76 *(0.40)* | 66.47 *(14.97)* | 3.48 *(1.30)* |
| | FLIP-IT | 5.27 *(1.3)* | 98.41 *(0.60)* | 79.33 *(10.93)* | 0.44 *(0.27)* | 99.98 *(0.02)* | 99.86 *(0.29)* | 2.94 *(1.3)* | 99.42 *(0.43)* | 84.62 *(15.14)* | 3.61 *(0.53)* | 99.15 *(0.19)* | 84.76 *(7.62)* | 3.06 *(0.80)* |
| | FLIP-MCL | 4.95 *(1.01)* | 98.11 *(0.50)* | 74.67 *(5.81)* | 0.54 *(0.22)* | 99.98 *(0.01)* | 100.00 *(0.00)* | 4.25 *(0.31)* | 99.07 *(0.17)* | 84.62 *(5.35)* | 2.31 *(0.46)* | 99.63 *(0.12)* | 92.28 *(3.37)* | 3.01 *(0.50)* |
| 5-shot | ViT (ECCV' 22) [4] | 3.42 | 98.60 | 95.00 | 1.98 | 99.75 | 94.00 | 2.31 | 99.75 | 87.69 | 7.34 | 97.77 | 66.90 | 3.76 |
| | ViTAF* (ECCV' 22) [4] | 2.92 | 99.62 | 91.66 | 1.40 | 99.92 | 98.57 | 1.64 | 99.64 | 91.53 | 5.39 | 98.67 | 76.05 | 3.31 |
| 5-shot | FLIP-V | 1.89 *(0.63)* | 99.67 *(0.13)* | 94.66 *(3.39)* | 1.01 *(0.67)* | 99.84 *(0.14)* | 96.56 *(5.48)* | 1.68 *(0.69)* | 99.47 *(0.38)* | 75.53 *(22.07)* | 2.27 *(0.65)* | 99.62 *(0.15)* | 93.23 *(5.42)* | 1.72 *(0.66)* |
| | FLIP-IT | 2.63 *(0.78)* | 99.55 *(0.10)* | 93.00 *(3.71)* | 0.46 *(0.29)* | 99.97 *(0.02)* | 99.86 *(0.29)* | 1.18 *(0.26)* | 99.83 *(0.06)* | 96.15 *(1.95)* | 3.07 *(0.55)* | 99.30 *(0.06)* | 83.15 *(3.00)* | 1.83 *(0.47)* |
| | FLIP-MCL | 3.42 *(0.16)* | 99.34 *(0.13)* | 82.67 *(7.35)* | 0.63 *(0.06)* | 99.98 *(0.01)* | 100.00 *(0.00)* | 1.52 *(0.09)* | 99.86 *(0.06)* | 97.23 *(1.04)* | 1.54 *(0.30)* | 99.81 *(0.06)* | 96.37 *(2.22)* | 1.77 *(0.15)* |

Table 2. Extending evaluation of cross-domain performance in Protocol 2 from 0-shot to 5-shot. We evaluate CASIA-SURF (**S**), CASIA-CeFA (**C**), and WMCA (**W**). We run each experiment 5 times under different seeds and report the mean HTER, AUC, and TPR@FPR=1%, along with their standard deviation (shown in brackets under the mean scores).

| | **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 (ECCV' 22) [4] | 7.98 | 97.97 | 73.61 | 11.13 | 95.46 | 47.59 | 13.35 | 94.13 | 49.97 | 10.82 |
| 0-shot | FLIP-V | 6.13 *(2.24)* | 97.84 *(1.54)* | 50.26 *(25.05)* | 10.89 *(1.93)* | 95.82 *(1.27)* | 53.93 *(8.27)* | 12.48 *(1.26)* | 94.43 *(0.97)* | 53.00 *(6.27)* | 9.83 *(1.35)* |
| | FLIP-IT | 4.89 *(0.85)* | 98.65 *(0.48)* | 59.14 *(14.63)* | 10.04 *(0.46)* | 96.48 *(0.56)* | 59.4 *(5.48)* | 15.68 *(0.89)* | 91.83 *(0.75)* | 43.27 *(5.93)* | 10.2 *(0.55)* |
| | FLIP-MCL | 4.46 *(1.10)* | 99.16 *(0.31)* | 83.86 *(6.62)* | 9.66 *(0.50)* | 96.69 *(0.87)* | 59.00 *(8.87)* | 11.71 *(0.45)* | 95.21 *(0.38)* | 57.98 *(2.18)* | 8.61 *(0.51)* |
| 5-shot | ViT (ECCV' 22) [4] | 4.30 | 99.16 | 83.55 | 7.69 | 97.66 | 68.33 | 12.26 | 94.40 | 42.59 | 6.06 |
| | ViTAF* (ECCV' 22) [4] | 2.91 | 99.71 | 92.65 | 6.00 | 98.55 | 78.56 | 11.60 | 95.03 | 60.12 | 5.12 |
| 5-shot | FLIP-V | 0.69 *(0.28)* | 99.96 *(0.05)* | 99.42 *(0.52)* | 3.68 *(1.32)* | 99.38 *(0.44)* | 85.87 *(7.05)* | 7.44 *(0.36)* | 97.62 *(0.27)* | 76.11 *(0.59)* | 2.95 *(0.49)* |
| | FLIP-IT | 0.80 *(0.44)* | 99.96 *(0.05)* | 98.67 *(1.40)* | 3.19 *(0.16)* | 99.44 *(0.11)* | 88.80 *(4.44)* | 7.63 *(0.60)* | 97.42 *(0.38)* | 71.6 *(3.49)* | 2.90 *(0.30)* |
| | FLIP-MCL | 2.43 *(0.78)* | 99.67 *(0.19)* | 95.16 *(2.4)* | 2.13 *(0.75)* | 99.74 *(0.13)* | 93.93 *(3.64)* | 6.2 *(0.53)* | 98.11 *(0.15)* | 79.44 *(1.29)* | 2.69 *(0.51)* |

Table 3. Extending evaluation of cross-domain performance in Protocol 3 from 0-shot to 5-shot. We evaluate all the 12 different combinations between MSU-MFSD (**M**), CASIA-MFSD (**C**), Replay Attack (**I**), and OULU-NPU (O). We run each experiment 5 times under different seeds and report the mean HTER along with their standard deviation (shown in brackets under the mean scores).

| | **Method** | **C → I** | **C → M** | **C → O** | **I → C** | **I → M** | **I → O** | **M → C** | **M → I** | **M → O** | **O → C** | **O → I** | **O → M** | **Avg.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0-shot | ADDA (CVPR' 17) [14] | 41.8 | 36.6 | - | 49.8 | 35.1 | - | 39.0 | 35.2 | - | - | - | - | 39.6 |
| | DRCN (ECCV' 16) [2] | 44.4 | 27.6 | - | 48.9 | 42.0 | - | 28.9 | 36.8 | - | - | - | - | 38.1 |
| | DupGAN (CVPR' 18) [3] | 42.4 | 33.4 | - | 46.5 | 36.2 | - | 27.1 | 35.4 | - | - | - | - | 36.8 |
| | KSA (TIFS' 18) [7] | 39.3 | 15.1 | - | 12.3 | 33.3 | - | 9.1 | 34.9 | - | - | - | - | 24.0 |
| | DR-UDA (TIFS' 20) [18] | 15.6 | 9.0 | 28.7 | 34.2 | 29.0 | 38.5 | 16.8 | 3.0 | 30.2 | 19.5 | 25.4 | 27.4 | 23.1 |
| | MDDR (CVPR' 20) [17] | 26.1 | 20.2 | 24.7 | 39.2 | 23.2 | 33.6 | 34.3 | 8.7 | 31.7 | 21.8 | 27.6 | 22.0 | 26.1 |
| | ADA (ICB' 19) [16] | 17.5 | 9.3 | 29.1 | 41.5 | 30.5 | 39.6 | 17.7 | 5.1 | 31.2 | 19.8 | 26.8 | 31.5 | 25.0 |
| | USDAN-Un (PR' 21) [6] | 16.0 | 9.2 | - | 30.2 | 25.8 | - | 13.3 | 3.4 | - | - | - | - | 16.3 |
| | GDA (ECCV' 22) [23] | 15.10 | 5.8 | - | 29.7 | 20.8 | - | 12.2 | 2.5 | - | - | - | - | 14.4 |
| | CDFTN-L (AAAI' 23) [22] | 1.7 | 8.1 | 29.9 | 11.9 | 9.6 | 29.9 | 8.8 | 1.3 | 25.6 | 19.1 | 5.8 | 6.3 | 13.2 |
| 0-shot | FLIP-V | 15.08 | 13.73 | 12.34 | 4.30 | 9.68 | 7.87 | 0.56 | 3.96 | 4.79 | 2.09 | 5.01 | 6.00 | 7.12 |
| | | *(4.60)* | *(4.81)* | *(4.41)* | *(2.41)* | *(1.62)* | *(1.39)* | *(0.46)* | *(0.77)* | *(0.98)* | *(0.63)* | *(1.41)* | *(1.69)* | *(2.10)* |
| | FLIP-IT | 12.33 | 15.18 | 7.98 | 1.12 | 8.37 | 6.98 | 0.19 | 5.21 | 4.96 | 0.16 | 4.27 | 5.63 | 6.03 |
| | | *(2.24)* | *(2.40)* | *(2.73)* | *(0.30)* | *(2.95)* | *(1.14)* | *(0.26)* | *(2.57)* | *(0.75)* | *(0.22)* | *(1.53)* | *(1.61)* | *(1.55)* |
| | FLIP-MCL | 10.57 | 7.15 | 3.91 | 0.68 | 7.22 | 4.22 | 0.19 | 5.88 | 3.95 | 0.19 | 5.69 | 8.40 | 4.84 |
| | | *(2.94)* | *(1.4)* | *(0.47)* | *(0.05)* | *(2.15)* | *(0.37)* | *(0.20)* | *(1.38)* | *(0.42)* | *(0.26)* | *(1.42)* | *(1.09)* | *(1.01)* |
| | ViTAF* | 4.98 | 4.38 | 10.85 | 2.55 | 5.08 | 8.63 | 1.59 | 1.79 | 7.92 | 1.65 | 3.4 | 4.4 | 4.77 |
| | | *(0.66)* | *(0.80)* | *(1.31)* | *(0.34)* | *(0.95)* | *(0.97)* | *(0.20)* | *(0.13)* | *(0.71)* | *(0.36)* | *(0.71)* | *(0.73)* | *(0.66)* |
| 5-shot | FLIP-V | 3.37 | 2.27 | 2.96 | 0.79 | 2.37 | 3.75 | 0.42 | 2.38 | 2.76 | 0.35 | 1.62 | 2.10 | 2.10 |
| | | *(1.23)* | *(1.19)* | *(0.68)* | *(0.26)* | *(1.26)* | *(0.92)* | *(0.30)* | *(0.34)* | *(0.47)* | *(0.29)* | *(0.34)* | *(0.68)* | *(0.66)* |
| | FLIP-IT | 4.11 | 5.22 | 4.20 | 0.42 | 2.22 | 3.20 | 0.40 | 2.31 | 3.21 | 0.16 | 2.45 | 3.78 | 2.64 |
| | | *(0.74)* | *(0.57)* | *(0.59)* | *(0.25)* | *(0.79)* | *(0.33)* | *(0.33)* | *(0.65)* | *(0.46)* | *(0.22)* | *(0.55)* | *(0.73)* | *(0.50)* |
| | FLIP-MCL | 4.18 | 5.27 | 2.48 | 0.65 | 3.68 | 2.56 | 0.19 | 1.74 | 2.43 | 0.23 | 2.58 | 4.10 | 2.51 |
| | | *(0.60)* | *(0.53)* | *(0.53)* | *(0.06)* | *(0.53)* | *(0.42)* | *(0.20)* | *(0.29)* | *(0.26)* | *(0.23)* | *(0.59)* | *(1.25)* | *(0.45)* |

Table 4. Computational complexity analysis for all the methods in the FLIP framework compared with the baseline methods.

| **Method** | **Training** | | | | **Inference** | | **Inference Time** |
|---|---|---|---|---|---|---|---|
| | **Image Encoder** | | **Text Encoder + Proj** | | | | |
| | **Parameters** | **FLOPs** | **Parameters** | **FLOPs** | **Parameters** | **FLOPs** | **(seconds/frame)** |
| ViT (ECCV' 22) | 86.19M | 17.58G | - | - | 86.19M | 17.58G | 0.007 |
| ViTAF* (ECCV' 22) | 92.02M | 18.68G | - | - | 92.02M | 18.68G | 0.020 |
| FLIP-V | 86.58M | 17.58G | - | - | 86.58M | 17.58G | 0.013 |
| FLIP-IT | 86.19M | 17.58G | 63.11M | 35.81G | 86.19M | 17.58G | 0.010 |
| FLIP-MCL | 86.19M | 52.74G | 83.05M | 35.86G | 86.19M | 17.58G | 0.010 |

## 3. Computational Complexity

We present the model size, training, and inference time computational complexity (computed on an NVIDIA Quadro RTX 6000) in Table 4. Kindly note that our image encoder (FLIP-V) is similar to [4] except that it is pre-trained using CLIP. However, FLIP-IT and FLIP-MCL require an additional text encoder during training. Furthermore, FLIP-MCL requires additional projection layers for the contrastive loss ($L_{simCLR}$). Thus, FLIP-IT and FLIP-MCL have some auxiliary parameters, only during training. Moreover, since FLIP-MCL requires three forward passes through the image encoder (original + 2 transformed views), it involves more computations. Once the training is complete, the embeddings for the text prompts can be pre-computed and stored. Hence, all the auxiliary parameters (text encoder + proj) can be discarded and only the image encoder is required for inference. Therefore, our inference time is similar to the baseline method [4], while our approach significantly improves the generalization to unseen domains.

## 4. Robustness to Unseen Spoof Type

To understand the robustness of the proposed FLIP-MCL method to unseen spoof types, we design an experiment to evaluate its performance, where the training and testing spoof types are completely different. We present the results in Table 5. Each dataset in Protocol 1 (M, C, I, O) contains real, print attack, and replay attack samples. We aggregate the samples of real, print, and replay from all 4 datasets and split each group into a train-test split of 80%-20%. For the *Replay* experiment, we train only on real and

Table 5. HTER performance on unseen spoof type at test time. *Replay* denotes training on real+print samples and testing on unseen replay samples. *Print* denotes training on real+replay samples and testing on unseen print samples.

| Method | *Replay* | *Print* |
|---|---|---|
| ViT (ECCV' 22) [4] | 4.69 | 10.36 |
| FLIP-MCL | 1.07 | 1.98 |

print samples and test on unseen replay samples. Similarly, we perform the *Print* experiment by training only on real and replay samples and testing on unseen print samples. We observe that for both the unseen testing scenarios (*Replay* & *Print*) the proposed FLIP-MCL method comfortably outperforms the baseline ViT thus demonstrating its generalizability. This validates the idea that aligning images to text descriptions can also handle unseen spoof types.

## References


[1] Zhihong Chen, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Feiyue Huang, and Xinyu Jin. Generalizable representation learning for mixture domain face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1132–1139, 2021. 2

[2] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 597–613. Springer, 2016. 3

[3] Lanqing Hu, Meina Kan, Shiguang Shan, and Xilin Chen. Duplex generative adversarial network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1498–1507, 2018. 3

[4] 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. 1, 2, 3, 4

[5] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020. 2

[6] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. *Pattern Recognition*, 115:107888, 2021. 3

[7] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(7):1794–1809, 2018. 3

[8] 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. 2

[9] Shice Liu, Shitao Lu, Hongyi Xu, Jing Yang, Shouhong Ding, and Lizhuang Ma. Feature generation and hypothesis verification for reliable face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1782–1791, 2022. 2

[10] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Mingwei Bi, Shouhong Ding, Jilin Li, Feiyue Huang, and Lizhuang Ma. Adaptive normalized representation learning for generalizable face anti-spoofing. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1469–1477, 2021. 2

[11] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Yuan Xie, and Lizhuang Ma. Dual reweighting domain generalization for face presentation attack detection. *arXiv preprint arXiv:2106.16128*, 2021. 2

[12] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10023–10031, 2019. 2

[13] Rui Shao, Xiangyuan Lan, and Pong C Yuen. Regularized fine-grained meta face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11974–11981, 2020. 2

[14] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 3

[15] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20281–20290, 2022. 2

[16] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Improving cross-database face presentation attack detection via adversarial domain adaptation. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019. 3

[17] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6678–6687, 2020. 2, 3

[18] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 16:56–69, 2020. 3

[19] Jingjing Wang, Jingyi Zhang, Ying Bian, Youyi Cai, Chunmao Wang, and Shiliang Pu. Self-domain adaptation for face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2746–2754, 2021. 2

[20] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4123–4133, 2022. 2

[21] Zitong Yu, Jun Wan, Yunxiao Qin, Xiaobai Li, Stan Z Li, and Guoying Zhao. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. *IEEE transactions on pattern analysis and machine intelligence*, 43(9):3005–3023, 2020. 2

[22] Haixiao Yue, Keyao Wang, Guosheng Zhang, Haocheng Feng, Junyu Han, Errui Ding, and Jingdong Wang. Cyclically disentangled feature translation for face anti-spoofing. *arXiv preprint arXiv:2212.03651*, 2022. 3

[23] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Kekai Sheng, Shouhong Ding, and Lizhuang Ma. Generative domain adaptation for face anti-spoofing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, pages 335–356. Springer, 2022. 2, 3