

Domain Agnostic Feature Learning for Image and Video Based Face Anti-spoofing : Supplementary Material

Suman Saha
ETH Zurich

suman.saha@vision.ee.ethz.ch

Wenhao Xu
ETH Zurich

wenhxu@student.ethz.ch

Menelaos Kanakis
ETH Zurich

menelaos.kanakis@vision.ee.ethz.ch

Stamatios Georgoulis
ETH Zurich

stamatios.georgoulis@vision.ee.ethz.ch

Yuhua Chen
ETH Zurich

yuhua.chen@vision.ee.ethz.ch

Danda Pani Paudel
ETH Zurich

paudel@vision.ee.ethz.ch

Luc Van Gool
KU Leuven & ETH Zurich

vangool@vision.ee.ethz.ch

1. Detailed network design

In this section, we present a detailed architectural design of the proposed network. We use the ResNet-50 as our backbone network and PyTorch for implementation purpose. In Table 1, we show the layer-wise network design of our proposed Class-conditional Domain Discriminator (CCDD) and the live/spoof classifier (LSC). Note that, the ResNet-50 backbone inputs a 224×224 image and outputs a 2048 dimensional feature vector which is passed as the input to both the CCDD and LSC. The FC31 and FC32 layers (i.e. the live and spoof heads) of CCDD output 3 softmax probability scores for the 3 source domains. The FC2 (or final) layer of the LSC outputs 2 softmax probability scores for the 2 class labels - “live” and “spoof”.

The shape of the input tensor to the LSTM network is $[T \times B \times 2048]$ where T is the sequence length and B is the SGD mini-batch size. We set T and B to 8 and 2 respectively. For each time step t , the LSTM outputs 256 dimensional feature vector, where $t = 1, 2, \dots, 8$. These 8 feature vectors are concatenated to a single feature vector of dim. 2048 which is then passed as input to the LSC.

2. Additional experimental details

We use four domain generalization datasets as in [9] which are created from the following publicly available face anti-spoofing datasets: Oulu-NPU [1] (O for short), CASIA-MFSD [13] (C for short), Idiap Replay-Attack [2] (I for short), and MSU-MFSD [12] (M for short). In Table 2, we present the training, validation and test set details for each of these four datasets. CASIA-MFSD and MSU-MFSD don't have validation sets and following the standard

practice, we use a subset of the training set as the validation set for both of these datasets. At inference time, we receive the predictions from both image- and video-based live/spoof classifiers (see Fig.3 in the main paper). As a final output, we select the one which gives the best performance on the validation set. We initialize the ResNet-50 backbone with ImageNet [4] pretrained weights. Please note, it is fair to compare our proposed video-based approach with the image-based baselines as all the datasets considered in this work are video-based, and thus, at the time of evaluation, the final classification score is generated per video, i.e. both image- and video-based methods follow the same standard video-based evaluation.

3. Experimental results with smaller backbone

In this section, we present our experimental results with a smaller backbone (i.e. ResNet-18) compared to the ResNet-50 used in the main paper. We replace the ResNet-50 backbone in our proposed framework with ResNet-18 and train the model. In Table 3, we show that even if our proposed framework uses a weaker backbone network, it shows consistent improvements on the four challenging domain generalization test sets. Note that, $I \& C \& M \rightarrow O$ has the smallest training set among these four domain generalization datasets (Table 2). For this smaller dataset, our proposed framework achieves better performance with a ResNet-18 backbone.

4. Assessing the model's generalizability

To assess the model's generalization ability, we increase the number of source domains. By doing so, we allow the

Table 1. The architectural details of the proposed network.

CCDD (Class-conditional Domain Discriminator)			Live/Spoof Classifier		
Layer	Input Dim.	Output Dim.	Layer	Input Dim.	Output Dim.
FC1	2048	1024	FC1	2048	512
ReLU			ReLU		
Dropout			Dropout		
FC2	1024	1024	FC2	512	2 (num. class labels)
ReLU					
Dropout					
FC31 (Live Head)	1024	3 (num. source domains)			
FC32 (Spoof Head)	1024	3 (num. source domains)			

Table 2. Domain generalization training, validation and test sets used in this work.

Dataset Name	Training Set	Validation Set	Test Set
O&C&I→M	Training sets from Oulu-NPU, CASIA-MFSD and Idiap Replay-Attack.	Validation sets from Oulu-NPU, CASIA-MFSD and Idiap Replay-Attack.	MSU-MFSD test set
O&M&I→C	Training sets from Oulu-NPU, MSU-MFSD and Idiap Replay-Attack.	Validation sets from Oulu-NPU, MSU-MFSD and Idiap Replay-Attack.	CASIA-MFSD test set
O&C&M→I	Training sets from Oulu-NPU, CASIA-MFSD and MSU-MFSD.	Validation sets from Oulu-NPU, CASIA-MFSD and MSU-MFSD.	Idiap Replay-Attack test set
I&C&M→O	Training sets from Idiap Replay-Attack, CASIA-MFSD and MSU-MFSD.	Validation sets from Idiap Replay-Attack, CASIA-MFSD and MSU-MFSD.	Oulu-NPU test set

Table 3. Face anti-spoofing performance (HTER%) improvements with smaller backbone network (ResNet-18).

Model	O&C&I→M	O&M&I→C	O&C&M→I	I&C&M→O
ResNet-18 backbone	27.5	31.67	21.63	14.83
Our model (uses ResNet-18 backbone)	22.5	28.52	20.38	12.78

Table 4. Assessing the model’s generalizability on three domain generalization test sets.

Model	S&O&I&R →C HTER(%)	S&O&C&R →I HTER(%)	S&C&I&R →O ACER(%)
ResNet-50	17.5	20.6	10.27
Our IB Network	14.0	14.7	8.05

network to see live and spoof examples with large variations in subjects, environmental conditions, attack instruments, video capturing devices, etc. One may argue that in this case the improvements merely come by adding more data. To ensure this is not the case, we also compare results against the ResNet baseline on the same data. As shown in Table 4, our imaged based network (DIB) achieves consistent improvements over the ResNet baseline, on three different domain generalization test settings. Note that, for the experiments in this section we used – only for training purposes – two more datasets, i.e. SiW (S for short) [7] and Idiap replay-mobile (R for short) [3]. Although, further improvements can be achieved by adding the video based network (LSTM and DVB), here we demonstrate improvements only in image-based FAS and exclude the video-based case. Following [7], when testing on Oulu-NPU dataset, we use the ACER metric and report results (in Table 4) by averaging the ACERs over the four test protocols.

5. Domain adaptation experiments

In this section, we compare the face anti-spoofing performance of our proposed approach with the existing domain adaptation based FAS methods [6, 10, 11]. For these experiments, we follow standard unsupervised domain adap-

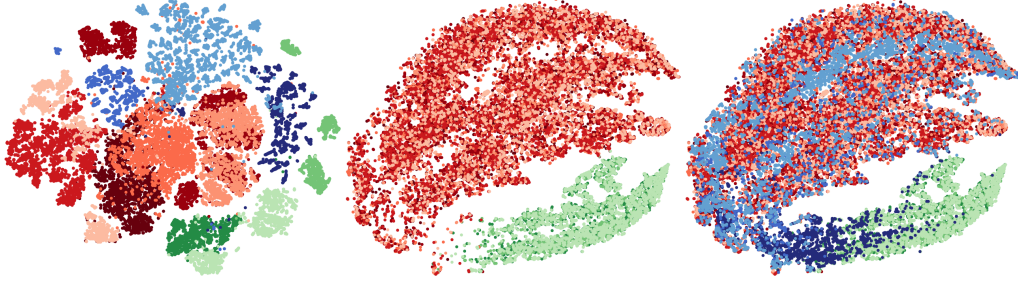
tation training setting, *i.e.* the network is trained using examples from a single source domain (with ground-truth labels) and unlabelled training examples from the target domain. To align with the domain adaptation training setup, we use the default domain discriminator (see Section 4.5, Fig.5). The results are presented in Table 5. Out of four domain adaptation test sets, our proposed framework outperforms [6, 10, 11] on the three test sets, and shows comparable results on the remaining one. These results demonstrate that our proposed model can be exploited under both domain adaptation and domain generalization settings. Note that, in this paper we are interested in the latter setting, yet as we observe from these experimental results, our model with small adaptations can achieve significant improvements for the former setting too.

Li *et al.* [6] have different strategies for domain adaptation and we select whatever gives out the best result for their model and compare to ours in Table 5. Li *et al.* [5] approach is not comparable to ours because they assume that different domains are just different camera models, which is quite restricting. In our case they are different datasets, allowing us to also address changes in spoofing mediums, illumination and background.

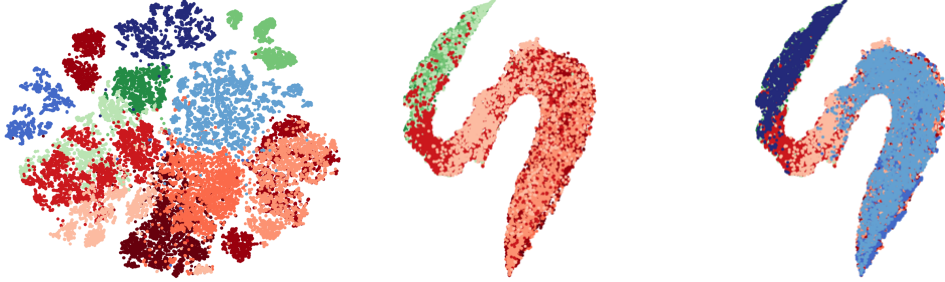
Table 5. Comparison to the existing domain adaptation based face anti-spoofing methods on four domain adaptation test sets.

Method	M→I HTER(%)	I→M HTER(%)	I→C HTER(%)	C→I HTER(%)
Li et al. [6]	33.30	33.20	12.30	39.20
Tu et al. [10]	27.50	25.83	-	-
Tu et al. [11]	25.80	23.50	23.50	21.40
Our full model	9.38	12.91	16.11	11.38

(a) O&C&I→M



(b) O&M&I→C



(c) I&C&M→O

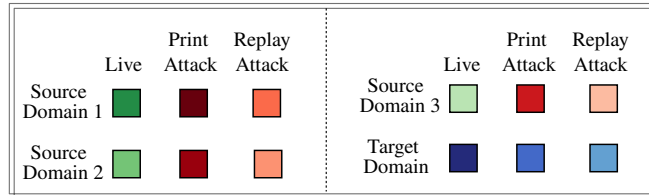
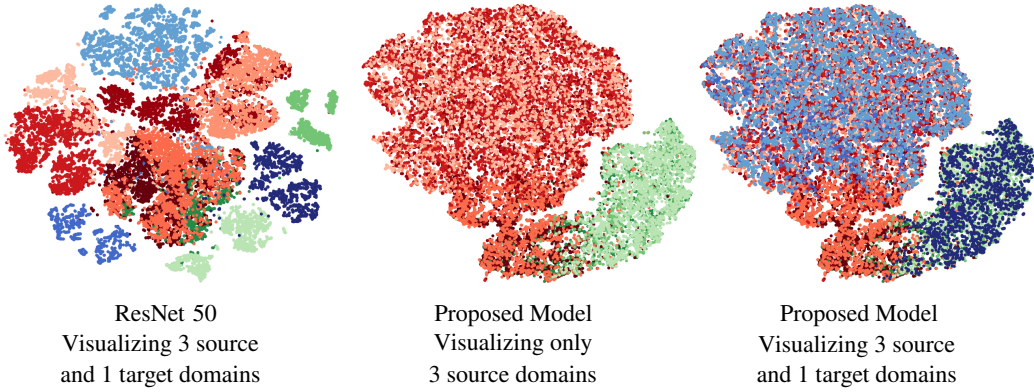


Figure 1. A t-SNE Visualization of the learned CNN features coming from ResNet-50 baseline and from our proposed network.

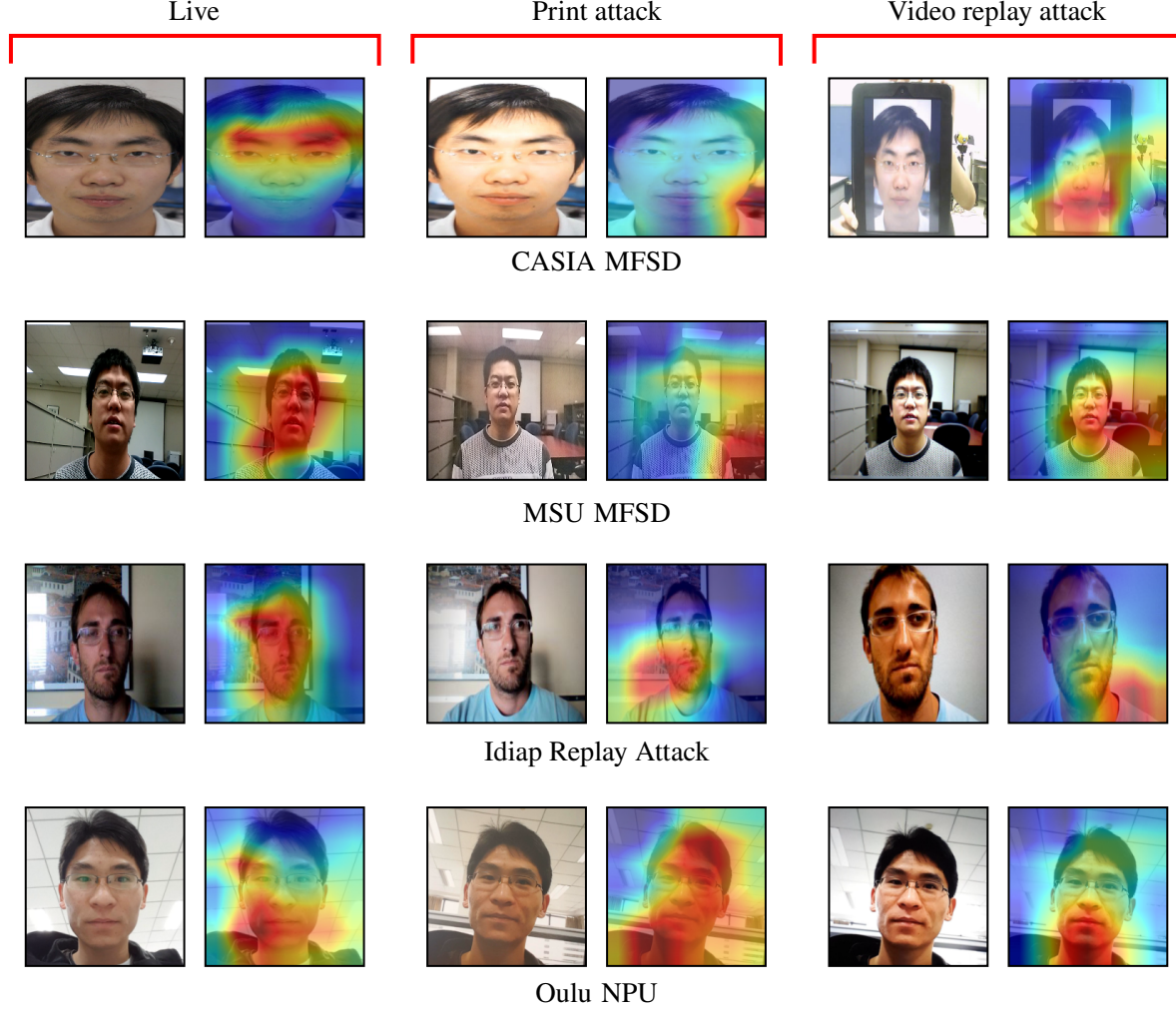


Figure 2. Class activation map visualization of the proposed network. For each column i.e. live, print attack and video replay attack, the original input images and their associated network class activation maps are shown.

6. t-SNE Visualization of the learned CNN features

We compared the t-SNE visualization (Section 4.4, Fig.4 in the main) of the CNN activations coming from the ResNet-50 baseline vs our proposed model which were trained and tested according to the domain generalization setup - $O \& C \& M \rightarrow I$ (see Table 2). In Fig. 1, we present a t-SNE visualization for the ResNet-50 vs our proposed model trained on the remaining three domain generalization training sets - $O \& C \& I \rightarrow M$, $O \& M \& I \rightarrow C$ and $I \& C \& M \rightarrow O$ (see Table 2). Each row in Fig. 1 represents a domain generalization train/test setup (see Table 2). The plots in the first column (in Fig. 1) are generated using ResNet-50 baseline model. Whereas, the plots in the second and third columns (in Fig. 1) are generated using our proposed model. For the

sake of better visualization, however, we have deactivated the visualization of the target domain in the second column.

Similar observations can be made as in Fig.4 in the main paper. The t-SNE plots in the second and third columns show that our model (1) learns more discriminative features for live and spoof images (second column); (2) aligns well the target domains live and spoof features with the source domains’ live and spoof features. In contrast, the ResNet-50 features show relatively weaker generalization ability on the target domain, as shown in Fig. 1 (first column).

7. Class activation map visualization

Similar to Section 4.6 in the main paper, here we present some additional class activation maps for the “live”, “print attack” and “replay attack” samples from the four face anti-spoofing datasets. Fig. 2 shows the class activation maps

which are generated using Grad-CAM [8]. Similar observations can be made (as in Section 4.6 in the main paper) from these activation maps, i.e. for “live” samples, the network activations are high around the facial regions. For “print attacks”, the network activations are high in the background regions (except Oulu-NPU), i.e. the network learns to classify print attacks by detecting the small artifacts often appears on the surface of the paper material (on which the face image was printed). The high resolution print attacks of Oulu-NPU might force the network to look at both facial regions as well as the background. For “video replay attack”, the network tries to gather information both from the facial regions and background. The important clues to classify a replay attack might include the moire patterns appears in the CRT displays, the unique texture of the display screen frame etc.

References

- [1] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 612–618. IEEE, 2017. [1](#)
- [2] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pages 1–7. IEEE, 2012. [1](#)
- [3] Artur Costa-Pazo, Sushil Bhattacharjee, Esteban Vazquez-Fernandez, and Sebastien Marcel. The replay-mobile face presentation-attack database. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7. IEEE, 2016. [2](#)
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. [1](#)
- [5] Haoliang Li, Peisong He, Shiqi Wang, Anderson Rocha, Xinghao Jiang, and Alex C Kot. Learning generalized deep feature representation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(10):2639–2652, 2018. [2](#)
- [6] 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. [2](#), [3](#)
- [7] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 389–398, 2018. [2](#)
- [8] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017. [5](#)
- [9] 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 Conference on Computer Vision and Pattern Recognition*, pages 10023–10031, 2019. [1](#)
- [10] Xiaoguang Tu, Hengsheng Zhang, Mei Xie, Yao Luo, Yuefei Zhang, and Zheng Ma. Deep transfer across domains for face anti-spoofing. *arXiv preprint arXiv:1901.05633*, 2019. [2](#), [3](#)
- [11] Xiaoguang Tu, Jian Zhao, Mei Xie, Guodong Du, Hengsheng Zhang, Jianshu Li, Zheng Ma, and Jiashi Feng. Learning generalizable and identity-discriminative representations for face anti-spoofing. *arXiv preprint arXiv:1901.05602*, 2019. [2](#), [3](#)
- [12] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015. [1](#)
- [13] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *ICB*, pages 26–31. IEEE, 2012. [1](#)