DOMAIN ADAPTATION FOR GENERALIZATION OF FACE PRESENTATION ATTACK DETECTION IN MOBILE SETTINGS WITH MINIMAL INFORMATION

Amir Mohammadi, Sushil Bhattacharjee, and Sébastien Marcel

Idiap Research Institute, Switzerland

ABSTRACT

With face-recognition (FR) increasingly replacing fingerprint sensors for user-authentication on mobile devices, presentation attacks (PA) have emerged as the single most significant hurdle for manufacturers of FR systems. Current machine-learning based presentation attack detection (PAD) systems, trained in a data-driven fashion, show excellent performance when evaluated in intra-dataset scenarios. Their performance typically degrades significantly in cross-dataset evaluations. This lack of generalization in current PAD systems makes them unsuitable for deployment in real-world scenarios. Considering each dataset as representing a different domain, domain adaptation techniques have been proposed as a solution to this generalization problem. Here, we propose a novel one class domain adaptation method which uses *domain guided pruning* to adapt a pre-trained PAD network to the target dataset. The proposed method works without the need of collecting PAs in the target domain (*i.e.*, with minimal information in the target domain). Experimental results on several datasets show promising performance improvements in cross-dataset evaluations.[1]

Index Terms— presentation attack detection, domain adaptation, domain generalization, pruning, feature selection

1. INTRODUCTION

Since the introduction of the Face ID by Apple on its iPhone X on the year 2017, mobile-phone manufacturers are increasingly turning to face-recognition (FR) as the technology of choice for user-authentication. Although modern, deep-learning based FR systems [1, 2, 3] achieve excellent recognition rates, they have also been shown to be highly vulnerable to *presentation attacks* (PA, also referred to as spoof-attacks) [4]. PAs are performed on the biometric sensor – the camera in an FR system. For example, the attacker may impersonate the identity of another person by presenting a printed face photo of the intended victim to the camera of the FR system. Such an attack is a PA, specifically a *print* PA. Another type of PA, called *replay* attack, is said to occur when the face image of the intended victim is presented using a digital display-device (such as a smart-phone screen, or tablet computer screen) to the camera. Non-PA face-biometric samples are called *bona fide* (BF) samples.

Countermeasures designed to prevent face-PAs are called face presentation attack detection (PAD) systems. Most modern face-PAD systems rely on machine-learning, and are trained to discriminate between BF and PA samples using large datasets. In recent years different research groups have publicly shared several datasets for face-PAD experiments [5, 6, 7]. Each dataset represents a specific *domain*. The term 'domain' encapsulates a broad range of parameters, including the cameras used to capture the biometric samples comprising the dataset, the environmental/imaging conditions, the various classes of PAs represented in the dataset, even the set of subjects providing the BF samples, and so on.

Face-PAD datasets include protocols defining mutually disjoint data subsets for *training*, *validation*, and *evaluation* of face-PAD methods. Typical face-PAD studies involve training a PAD method using the training set of a specific dataset, tuning hyper-parameters (if any) of the training method using the validation set, and finally classifying the samples in the evaluation set using the trained face-PAD system, to quantify the performance of the system in an unbiased fashion.

For a face-PAD system trained using a given dataset (the *source* domain), two evaluation scenarios are possible:
1. *intra-domain* evaluation: the evaluation set is taken from the same source dataset, or,
2. *cross-domain* evaluation: the evaluation set comes from another, *target* dataset, different from the source dataset.
Current face-PAD systems, especially those based on convolutional neural networks (CNN), show promising PAD performance in intra-domain evaluation scenarios [8]. Typically, however, their performance degrades significantly when tested in cross-domain scenarios [8]. This lack of generalization may be attributed to the *domain shift* (also called *covariate shift* or *dataset bias*) present between *source* and *target* datasets [9, 10].

Domain shift in face-PAD datasets is illustrated in Figure 1, which shows face-image samples of various classes from different datasets. We note from the figure that the samples from each class can change drastically between datasets. Among face-PAD datasets, domain shift may be caused by many factors, such as the camera device, resolution of images, distance of the subject to the camera, the instrument used to create the attack, lighting conditions, and identity. *Domain adaptation* and *domain generalization* [9, 10] methods have been developed to mitigate the problem of domain-shift in machine-learning.

For heterogeneous face recognition [11] and speech recognition [12], it has been shown that adapting only a few layers of a CNN to a *target* dataset can significantly improve the per-

This work was funded by the Research Council of Norway (Grant No. IKTPLUSS 248030/O70). Amir Mohammadi (amir.mohammadi@idiap.ch) is also affiliated with École Polytechnique Fédérale de Lausanne (EPFL).

[1] Source code: https://gitlab.idiap.ch/bob/bob.paper.icassp2020_domain_guided_pruning

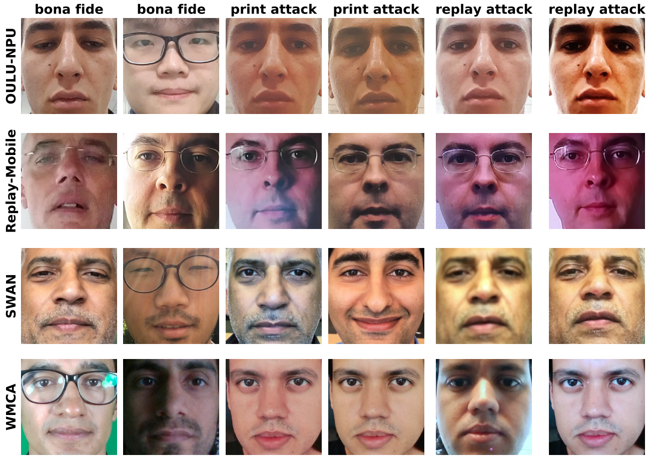

Fig. 1: Examples of domain shift between datasets in face PAD. The samples of each row belong to the same dataset and samples of each column are from the same class. The samples can be very different within classes while they can be very similar between classes. Also, samples from each class can change drastically between datasets.

formance of the recognition system in the target domain. Both studies showed that adapting a small number of initial layers (layers closest to the input layer) led to the most significant performance improvements in the target domain. In these CNNs the initial layers may be considered *domain specific*, whereas the remaining layers may be considered *domain invariant* and more task specific [11]. The approaches proposed in [11, 12] have been developed under the assumptions that the source and target domains are clearly defined. For the problem of face-PAD, we are often not be able to clearly identify distinct domains.

A more significant problem is that of data collection in the target domain. Specifically, whereas BF samples may be collected in the target domain at reasonable cost, collecting PAs in the target domain is usually much more expensive, if not impossible. Also, in real-world scenarios, a PAD system may be presented with attacks of previously unseen classes of PA.

In this work, we propose a novel domain adaptation method relying on minimal information – only BF samples from the target domain. We hypothesize that, in a CNN trained for PAD using a *source* dataset, some learned filters in a layer are domain specific and others are domain invariant. We assume that by pruning domain specific layers, which do not generalize to the target dataset, we can improve the performance of the model on the target dataset.

In the next section we discuss related works on domain adaptation for face-PAD. The proposed domain adaptation method is presented in Section 3. Implementation details are outlined in Section 4. Experimental results are presented in Section 5 and conclusions are summarized in Section 6.

2. RELATED WORK

Of the many domain adaptation and domain generalization methods [13], and many developed for deep learning based models [10], some of these methods have been applied to the problem of face PAD [14, 15, 16, 17]. In most methods, the distribution of source and target features are matched in a learned feature space. If the features have similar distributions, a classifier trained on features of the source samples can also be used to classify the target samples.

Li *et al.* [14] propose a domain generalization method by training on multiple domains and testing on an unseen domain. They consider different *camera devices* used to record face videos as different *domains*. To conduct the experiments, three publicly available face PAD datasets (REPLAY ATTACK [18], MSU MFSD [19], and CASIA FASD [20]) are combined to create 10 protocols. In each protocol, data from one camera is set aside as the target domain, and a subset of the remaining cameras are used as source domains. Maximum Mean Discrepancy [21] (MMD) is used to match the distribution of features at training. The authors report that the addition of the domain generalization method in the proposed 3D CNN brings 10% of absolute improvement in half total error rate (HTER) on average. The main drawback of this method is that, during training, the factors (here, camera device) that cause domain shift between datasets must be known beforehand and labeled. In other scenarios it may not be possible to explicitly identify the factors causing domain-shift. Also, some factors, such as lighting conditions, are more difficult to categorize and label.

Li *et al.* [15] also introduce an unsupervised domain adaptation method based on MMD where each dataset is considered a domain. The objective is to learn a mapping that brings the distribution of source features close to the distribution of target features in a reproducing kernel Hilbert space (RKHS). The mapping is learned in a way to minimize MMD between source and target distributions in the RKHS. Once the source features are mapped, a two-class classifier is trained on the mapped source features. To classify a new sample in the target domain, target features in the kernel space are labeled using the trained classifier. The authors evaluate the proposed method on several hand-crafted and also deep learning based features. For deep learning based features, embeddings extracted from AlexNet [22] are used. They compare the cross-domain evaluation results with and without the proposed domain adaptation method, and report an absolute improvement of 24% in HTER (on average, using AlexNet embeddings) when data from both BF and PAs are used to compute MMD in an unsupervised manner. When only BF data is used to compute MMD, the absolute improvement in HTER on average is reported as 14%. The datasets that are used for evaluation are the same as in [14]. In this approach, to classify a new sample, the target domain must be known *a priori* as different feature extractors are to be used for different domains, and no feature extractor can be designed for samples from unknown domains.

In [16], fast style transfer [23] is used for domain adaptation. The method works by training a CNN for PAD in the source domain and also simultaneously learning the *style* of source samples using fast style transfer. At test time, given a face image, the style of the image is first transferred to the source domain and then it is given to the CNN for classification. They report an absolute improvement of 2.9% in HTER on average in cross-domain evaluations when the domain adaptation method was added. The datasets that are used for evaluation are the same as in [14]. This method improves the performance of the PAD system only marginally and there are also some problems when the source domain images have a lower resolution than target domain images.

Shao *et al.* [17] present an approach to domain generalization by training on multiple domains (here, PAD datasets) and testing on an unseen domain. Using adversarial domain adaptation [24], they train a feature extractor, which is a CNN, that outputs features that are both useful for PAD and are also domain invariant. In other words, the features from a domain are indistinguishable from other domains. Four PAD datasets of OULU-NPU [6], REPLAY ATTACK, MSU MFSD, and CASIA FASD are used and an average absolute improvement of 12% in HTER is reported when the additional domain generalization method is added. Again, this work requires samples from both BF and PA classes.

3. DOMAIN GUIDED PRUNING

In this work we hypothesize that, in a CNN trained for PAD, some of the filters learned in the initial layers (layers closer to input) are robust filters and generalize well to the *target* dataset whereas others are more specific to the *source* dataset. We further hypothesize that by pruning the filters that do not generalize well from one dataset to another, the performance of the network on the target dataset can be improved.

One way of quantifying domain shift at a given layer in a CNN is by computing a *feature divergence* measure (FDM) [25]. Given two datasets representing different domains, A and B, we want to determine, how often, on average, a specific filter in layer, L, is activated in each domain. Let us denote the average value of a filter over the spatial dimensions as f and assume a Gaussian distribution for f with mean μ and variance σ^2. The symmetric Kullback-Leibler (KL) divergence of this filter between domains A and B is:

$$D(f_A||f_B)=KL(f_A||f_B)+KL(f_B||f_A) \quad (1)$$

where $KL(f_A||f_B)$ is the KL divergence of two Gaussian distributions [26, 27]. Let us denote $D(f_{iA}||f_{iB})$ as the symmetric KL divergence of the i^{th} filter in layer L. Then, the average feature divergence of layer L is given by $D(L_A||L_B)=\frac{1}{C}\sum_{i=1}^{C}D(f_{iA}||f_{iB})$ where C is the total number of filters in layer L. Higher values in Equation 1 indicate that the given filter is activated differently between datasets. Thus, the FDM for a given filter indicates whether it sensitive to the domain shift.

It is not always feasible to capture PAs in the target domain. Hence, we propose a method that relies on only BF samples from the target domain. The details of the proposed method are as follows. Assuming that there exist two datasets that represents different domains, A (source) and B (target), and the CNN model is trained on the source (A) domain:

1. Compute FDM (Eqn. 1) for each filter F at the layer L using only BF samples of the *training set* of datasets A and B.
2. Prune N percent of the filters[2] of layer L which contribute to the most feature divergence values at layer L.
3. Re-train the layers $L+1$ and after on the *training set* of the *source* dataset again (not the *target* dataset since it is assumed that no PAs are available for training in the target dataset) using the same classification loss-function to account for the pruned filters.

The pruned CNN is evaluated on the *evaluation set* of the target dataset. Intuitively, this method works like a feature selection method. The first L layers following the input layer of the CNN may be seen as a *feature extractor*. Layers $L+1$ and after may be seen as a *classifier*. Then, by pruning *features* at layer L and retraining the classifier, the classifier is limited to use only robust features for prediction.

4. IMPLEMENTATION DETAILS

Table 1: Details of DeepPixBiS. F is the number of filters and S is the stride. Layers 1 to 5 are identical with DenseNet-161 [28]. The input to the network is a 224×224 pixel color face image.

#	Layer	Details	Output Shape	Number of Parameters
1	conv0	Conv2D F=7 S=2	112 x 112 x 96	14,496
	pool0	MaxPool2D F=3 S=2	56 x 56 x 96	0
2	dense1	Dense Block	56 x 56 x 384	756,288
3	trans1	Transition Block	28 x 28 x 192	75,264
4	dense2	Dense Block	28 x 28 x 768	2,077,056
5	trans2	Transition Block	14 x 14 x 384	297,984
6	dec	Conv2D F=1 S=1	14 x 14 x 1	385

The face PAD datasets used in this study are: OULU-NPU [6], Replay-Mobile [5], SWAN [29], and WMCA [7]. Only print and replay attacks are considered from each dataset. OULU-NPU has been chosen as the *source* dataset in our experiments and other three datasets are considered *target* datasets. Since the proposed method uses only BF samples, we have also tested a scenario when the model is not pruned using BF samples of a PAD dataset but rather, using an FR dataset. We have used the still images of IARPA Janus Benchmark C (IJB-C)[3]FR dataset for this purpose. We removed the low quality face images of this dataset manually and have used around 3000 high quality face images. The classification performances is reported in terms of area under the curve (AUC) of *log-scale* receiver operating characteristic (ROC) curves. The ROC curves are computed with false positive rate ($APCER$ in [30]) along the x-axis (log-scale) and true positive rate ($1-BPCER$ in [30]) on the y-axis[4]. The proposed method is tested on the DeepPixBiS CNN architecture [8] which is detailed in Table 1.

5. EXPERIMENTS

For analysis, feature divergences of each layer of DeepPixBiS between the *training set* of OULU-NPU and the *evaluation set* of four datasets are shown in Figure 2. We can see that the last three final layers have the highest divergences compared to the three initial layers. We assume that re-training these last three layers, using only robust filters of the fourth layer from the end, would improve the performance. Hence, in our experiments, the *trans1* layer is chosen as the layer (L) at which the filters are pruned, and layers 4 to 6 are re-trained after pruning.

The proposed method requires the feature divergences for each filter at layer L to be computed between the source and target datasets. We have computed the feature divergences for

[2] Pruning can be implemented either by multiplying the output of a filter by zero, or by removing the filter entirely from calculations to reduce the computational cost. Both methods result in the same behavior.

[3] `https://www.nist.gov/programs-projects/face-challenges`

[4] Note that since AUC of *log-scale* ROCs are reported, their values can be higher than 1.

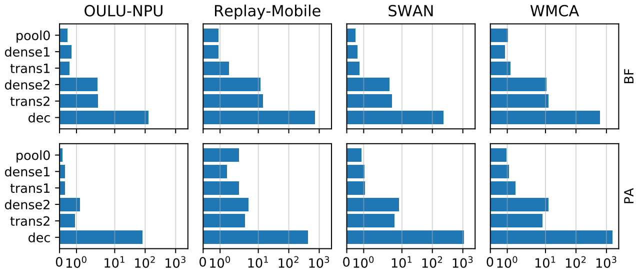

Fig. 2: Feature divergence at different layers computed between the *training set* of OULU-NPU and the *evaluation set* of OULU-NPU, Replay-Mobile, SWAN, and WMCA. FDM values are computed per layer (y-axis) and per class (top row for BF and bottom row for PA).

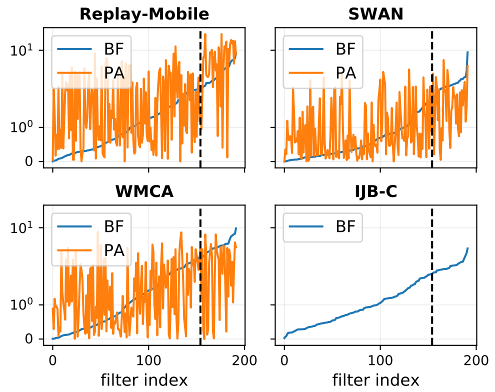

Fig. 3: FDM of each filter at layer L between OULU-NPU (source) and four target datasets: Replay-Mobile, SWAN, WMCA, and IJB-C.

both BF and PAs. Note that FDM values for PAs have been computed only for illustrative purposes – they are not used in the pruning process. We have observed that some of the pruned filters also have high FDM values for PAs. FDM values between the *training set* of OULU-NPU and the *training set* of four datasets are shown in Figure 3. In each plot, the filter indices are sorted in ascending order of FDM values calculated using the BF samples. The index at which $N\%$ of the filters would be pruned is shown using a vertical dashed black line. N was intuitively chosen to be 20 by observing the feature divergences in the plots.

The proposed method is evaluated using the following approach. First, DeepPixBiS is trained on OULU-NPU dataset and is evaluated on all four PAD datasets. This establishes *intra-domain* and *cross-domain* evaluations of the baseline. Then, DeepPixBiS is pruned according to the proposed method using four datasets; The resulting CNN is tested again on all four PAD datasets. The results are shown in Figure 4. The dataset that is used for pruning is mentioned in the plot-title. We can draw several conclusions from this figure:

1. Pruning does not significantly affect the performance of the model on the *source* dataset. This is shown in the results by testing all models on the OULU-NPU dataset.

2. We observe the performance improvement between the baseline and the proposed domain guided pruning method when the both test dataset is used for pruning. For example, comparing the performance of the original DeepPixBiS model on the Replay-Mobile dataset with that of the new version of the DeepPixBiS model pruned using the Replay-Mobile dataset, we note that the new model performs slightly better on the target (Replay-Mobile) dataset. We note that a model pruned using the target dataset performs better in the target domain. This improvement is clear for the WMCA dataset (AUC increases from $\sim$0.75 to $\sim$1.1).

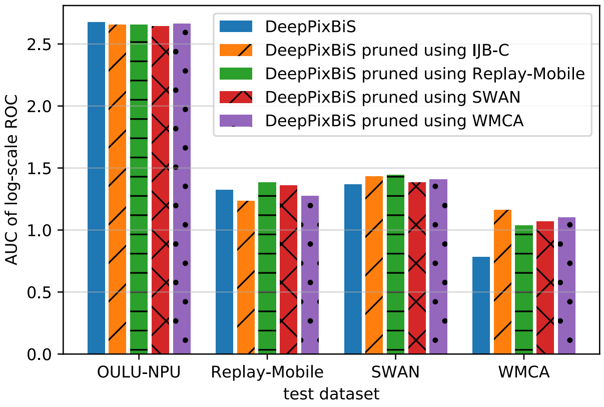

Fig. 4: Performance evaluation of the proposed method. The higher the value the better is the performance of the system. The evaluation-dataset is mentioned on the x axis. The models are compared to the baseline when no pruning is performed.

3. The effect of pruning using a different dataset than the test dataset is also observed. For example, we note that when the DeepPixBiS architecture is pruned using the IJB-C dataset, its performance degrades slightly on the Replay-Mobile dataset, improves slightly on the SWAN dataset, and improves significantly on the WMCA dataset.

6. CONCLUSIONS

In this work we have formulated the problem of generalization in PAD systems as a domain adaptation problem. Domain adaptation methods are often designed assuming that sufficient training data is available for all classes in the target domain. In reality, although collecting new BF samples in a target domain is usually affordable, collecting presentation attacks in the target domain may be quite expensive (and impossible for *unseen* attacks).

Here we have proposed a domain adaptation method, based on *domain guided pruning* of CNNs. The proposed method requires only BF presentations in the target domain. We present experimental results on three target datasets: Replay-Mobile, SWAN, and WMCA. These results lead to the following conclusions:

(a) When the CNN is pruned using the target dataset, the performance of the pruned model increases on the target dataset. Specifically, the pruned CNN performed significantly better than the baseline CNN, on the WMCA dataset.

(b) Pruning did not degrade the performance of the model on the source dataset.

These results give us confidence that the proposed pruning method is applicable to *intra-domain* and *cross-domain* evaluation scenarios.

We have also demonstrated that the proposed method can also be implemented as a *domain generalization* method. This is done by pruning the CNN using BF samples of an FR dataset before evaluating the pruned CNN on unseen PAD datasets.

7. REFERENCES

[1] David Sandberg, "Facenet: Face recognition using tensorflow," 2017.

[2] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015, vol. 1, p. 6.

[3] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan, "A light CNN for deep face representation with noisy labels," *arXiv preprint arXiv:1511.02683*, 2015.

[4] Amir Mohammadi, Sushil Bhattacharjee, and Sébastien Marcel, "Deeply vulnerable: A study of the robustness of face recognition to presentation attacks," *IET Biometrics*, vol. 7, no. 1, pp. 15–26, 2017.

[5] Artur Costa-Pazo, Sushil Bhattacharjee, Esteban Vazquez-Fernandez, and Sebastien Marcel, "The REPLAY-MOBILE Face Presentation-Attack Database," in *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of The*. 2016, pp. 1–7, IEEE.

[6] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference On*. 2017, pp. 612–618, IEEE.

[7] 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*, 2019.

[8] Anjith George and Sébastien Marcel, "Deep Pixel-wise Binary Supervision for Face Presentation Attack Detection," in *International Conference on Biometrics*, 2019.

[9] A. Gretton, AJ. Smola, J. Huang, M. Schmittfull, KM. Borgwardt, and B. Schölkopf, "Covariate shift and local learning by distribution matching," in *Dataset Shift in Machine Learning*, pp. 131–160. Biologische Kybernetik, Cambridge, MA, USA, 2009.

[10] Mei Wang and Weihong Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.

[11] T. de Freitas Pereira, A. Anjos, and S. Marcel, "Heterogeneous Face Recognition Using Domain Specific Units," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1803–1816, July 2019.

[12] Joel Shor, Dotan Emanuel, Oran Lang, Omry Tuval, Michael Brenner, Julie Cattiau, Fernando Vieira, Maeve McNally, Taylor Charbonneau, and Melissa Nollstadt, "Personalizing ASR for Dysarthric and Accented Speech with Limited Data," *arXiv preprint arXiv:1907.13511*, 2019.

[13] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual Domain Adaptation: A survey of recent advances," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 53–69, May 2015.

[14] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot, "Learning Generalized Deep Feature Representation for Face Anti-Spoofing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2639–2652, Oct. 2018.

[15] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1794–1809, 2018.

[16] 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.

[17] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C. Yuen, "Multi-Adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[18] Ivana Chingovska, André Anjos, and Sébastien Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of The*. 2012, pp. 1–7, IEEE.

[19] D. Wen, H. Han, and A. K. Jain, "Face Spoof Detection With Image Distortion Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[20] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z. Li, "A face antispoofing database with diverse attacks," in *Biometrics (ICB), 2012 5th IAPR International Conference On*. 2012, pp. 26–31, IEEE.

[21] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. Mar, pp. 723–773, 2012.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[23] Logan Engstrom, "Fast style transfer," 2016, commit xxxxxxx.

[24] 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*, 2017, pp. 7167–7176.

[25] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang, "Two at once: Enhancing learning and generalization capacities via ibn-net," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 464–479.

[26] Solomon Kullback and Richard A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[27] Solomon Kullback, *Information Theory and Statistics*, Courier Corporation, 1997.

[28] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[29] Raghavendra Ramachandra, Martin Stokkenes, Amir Mohammadi, Sushma Venkatesh, Kiran Raja, Pankaj Wasnik, Eric Poiret, Sébastien Marcel, and Christoph Busch, "Smartphone Multi-modal Biometric Authentication: Database and Evaluation," *arXiv:1912.02487 [cs]*, Dec. 2019.

[30] "ISO/IEC DIS 30107-3. Information Technology – Biometric presentation attack detection – Part 3: Testing and reporting," Standard, International Organization for Standardization, Geneva, CH, Jan. 2016.