Spoof detection using x-vector and feature switching

5 authors, including:

Mari Ganesh Kumar
Indian Institute of Technology Madras
8 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Saranya M S
Indian Institute of Technology Madras
11 PUBLICATIONS **12** CITATIONS

SEE PROFILE

Hema A Murthy
Indian Institute of Technology Madras
176 PUBLICATIONS **2,184** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Text-To-Speech View project

Text-to-speech-synthesis View project

Spoof detection using *x*-vector and feature switching

Mari Ganesh Kumar[1], *Suvidha Rupesh Kumar*[2], *Saranya M S*[1], *B. Bharathi*[2], *Hema A. Murthy*[1]

[1]Indian Institute of Technology Madras
[2]SSN college of Engineering

{mari,saranms,hema}@cse.iitm.ac.in, {suvidhark,bharathib}@ssn.edu.in

Abstract

Detecting spoofed utterances is a fundamental problem in voice-based biometrics. Spoofing can be performed either by logical accesses like speech synthesis, voice conversion or by physical accesses such as replaying the pre-recorded utterance. Inspired by the state-of-the-art *x*-vector based speaker verification approach, this paper proposes a deep neural network (DNN) architecture for spoof detection from both logical and physical access. A novelty of the *x*-vector approach vis-a-vis conventional DNN based systems is that it can handle variable length utterances during testing. Performance of the proposed x-vector systems and the baseline Gaussian mixture model (GMM) systems is analyzed on the ASV-spoof-2019 dataset. The proposed system surpasses the GMM system for physical access, whereas the GMM system detects logical access better. Compared to the GMM systems, the proposed *x*-vector approach gives an average relative improvement of 14.64% across four features for physical access, in terms of minimum tandem cost detection function (min-t-DCF). When combined with the decision-level feature switching (DLFS) paradigm, the best system in the proposed approach outperforms the best baseline systems with a relative improvement of 73.17% and 40.04% for both logical and physical access, respectively.

Index Terms: anti-spoofing, voice-biometrics, GMM, x-vectors

1. Introduction

Although automatic speaker verification (ASV) systems are robust to impostor threats [1] and acoustic variations, they are vulnerable when subjected to presentation attacks. Presenting a fake biometric sample to a biometric detection system is a presentation attack[1]. The process of this deliberate evasion is called spoofing. Spoofing at sample acquisition stage can be classified into two categories namely, logical access (LA) and physical access (PA) [2]. Synthesizing spoofing samples with speech synthesis (SS) or voice conversion (VC) approach are categorized as LA while replaying a pre-recorded original audio sample to access the verification system falls under the PA category. The primary objective of ASV-spoof-challenge proposed in 2015 was to detect logical access. Since the implementation of PA is easier than LA, the former attack is a greater threat than later. ASV-spoof-challenge in 2017 focused on identifying physical access. Numerous spoof detection algorithms have been proposed since then for both LA [3–5] and PA [6–8].

ASV-spoof-2019 challenge conducted this year focused on detecting spoofed utterances synthesized by both LA and PA. Unlike the previous anti-spoofing challenges, equal error rate (EER) was not used as the evaluation metric due to its ill-suited operating point for user applications like telephone banking [9].

Hence a new metric termed as a minimum normalized tandem detection cost function (min-t-DCF) is provided as the evaluation metric. The min-t-DCF considers the false alarms and misses for both countermeasure system as well as the ASV system, along with the prior probabilities of target and spoof trials. The details of min-t-DCF is discussed in [9, 10]. Scores from a *x*-vector based speaker verification system are used along with the statistics of the spoof detection system to estimate min-t-DCF. *x*-vector is a DNN based state-of-the-art speaker verification technique that embeds the speaker characteristics in low-dimensional fixed-length vectors from variable length utterances.

In this paper, inspired by the *x*-vector based ASV system, we propose a similar spoof detection system for identifying both logical and physical access. To implement the *x*-vector system for spoof detection, the following changes are made to the neural network architecture proposed in [11]: (i) The last layer in the ASV system's architecture is modified to handle the two-class problem of spoof detection. (ii) Instead of the standard cross-entropy loss function, a new focal loss function [12] is used to give more focus on hard and misclassified examples. The proposed *x*-vector classifier outperforms the baseline GMM classifier for physical access while GMM classifier outruns the *x*-vector system in detecting logical access. Owing to the success of decision-level feature switching (DLFS) system on ASV-spoof-2017 dataset [8], the same is used here to exploit the property of different features in capturing different kinds of spoofing conditions. The focus of this paper is three-fold: Firstly, a comparison of four different systems submitted to ASV-spoof-2019 challenge is discussed. Secondly, we propose the novel *x*-vector based spoof detection system. Finally, by using DLFS on individual feature system, the performance is further improved.

The rest of the paper is organized as follows: Section 2 discusses the details of spoof detection approaches in the literature. A brief description of ASV-spoof-2019 dataset is given in Section 3. Section 4 gives a brief overview of the *x*-vector based ASV system. The proposed *x*-vector architecture for spoof detection is explained in Section 5. Section 6 discusses the details of baseline GMM systems, the proposed *x*-vector systems, and the DLFS systems. A comprehensive analysis on the performance of various systems is given in Section 7 followed by the conclusion in Section 8.

2. Prior works on spoof detection

The ASV-spoof-2015 challenge targeted ten different types of logical access [13]. A combination of auditory transformation based on cochlear filter cepstral coefficients (CFCC) and instantaneous frequency (IF) termed as CFCCIF is proposed as the best feature to detect these LAs in [5]. Score fusion of CFCCIF and MFCC was adjudged as the best system with an average

---

[1]https://www.iso.org/standard/53227.html

EER of 1.2% across all the ten conditions. Various LA spoof detection systems submitted to the challenge are detailed in [4].

The speech corpus used in ASV-spoof-2017 challenge has the spoofed instances generated by recording and replaying the bonafide trials of speakers in different environments(E) using various recording(R) and playback devices(P). Physical attack is harder than logical access as the spoofed utterance of a bonafide trial may come from various E-R-P combinations. The evaluation subset of the ASV-spoof-2017 dataset tried to simulate this 'in-wild' condition by generating the spoofed instances from different E-R-P combinations. A light convolutional neural network (CNN) [14] system outperformed all other systems submitted to the challenge. In [7] an end-to-end neural network (NN) with attention masking was proposed to learn the difference in the spectrogram of bonafide and the replayed utterances. This end-to-end attention masking system pre-trained on ImageNet dataset [15] gives an ideal performance with zero percent EER. DLFS paradigm proposed in [8], uses information from multiple feature spaces. This technique outperforms all other replay attack detection systems in the literature except the ideal NN system with zero percent EER.

3. Dataset Description

Similar to the ASV-spoof-2015 and ASV-spoof-2017 corpus, [16] ASV-spoof-2019 also has three subsets namely, training (train), development (dev), and evaluation (eval). Different subsets of data are used for LA and PA attacks. The duration of each utterance is approximately two seconds. Unlike the "in-wild" spoofed trials of the ASV-spoof-2017 corpus, in this dataset, the spoofed trials for physical access are generated in controlled acoustic conditions [17]. The latest best performing text-to-speech synthesis and voice conversion algorithms are used to generate the spoofed trials for logical access category. These algorithms are better than the algorithms used in ASV-spoof-2015. The number of trials in each subset is listed in the Table 1. The number of trials in evaluation subsets of LA and PA are 71,747 and 137,457 respectively. The metadata of the evaluation, as well as the ground truth, are yet to be released.

Table 1: *Number of trials in development and training subsets*

Attack	Subsets	No. of speakers		No. of trials	
		Male	Female	Bonafide	Spoofed
LA	train	8	12	2580	22800
	dev	8	12	2548	22296
PA	train	8	12	5400	48600
	dev	8	12	5400	24300

4. *x*-vectors in speaker recognition

i-vectors were the state-of-the-approach for text-independent speaker recognition since 2010 [18]. An alternate approach proposed in [19] extracts DNN embeddings termed as *x*-vectors from a NN using a temporal pooling layer. This pooling layer facilitates the NN to discriminate the speakers from variable-length input speech segments. During testing, the fixed dimensional *x*-vectors are extracted and are compared with the training data embeddings using some scoring approach.

Speaker embeddings are extracted in [19] from variable length acoustic segments using a DNN with a multi-class cross-entropy loss function. The DNN consists of few time delay
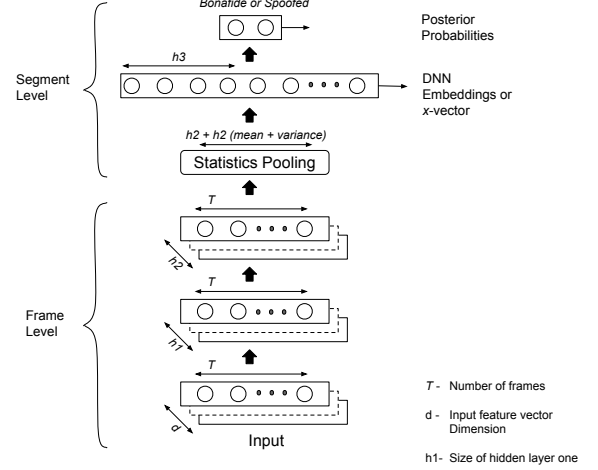

Figure 1: *Modified* x-*vector architecture for spoof detection*

neural network (TDNN) layers to enhance frame-level representation. A pooling layer aggregates the frame-level representations, followed by few additional layers to handle segment-level representations. Finally, a softmax layer to get posterior probabilities of each speaker. This approach mainly aims (i) to produce the speaker embeddings at utterance level rather than frame level and (ii) to generalize well, to handle the unseen speakers. The main advantage of this *x*-vector architecture is to handle the short duration utterances. *x*-vector results in [19] are shown to outperform the i-vector systems for short utterances of duration less than 10 seconds.

5. *x*-vectors for spoof detection

Similar to speaker characteristics, the signatures of the spoofing techniques will be present in the entire utterance. In this work, since the *x*-vectors capture the utterance level information, the ASV *x*-vector briefed in Section 4 is modified to detect spoofed utterances. ASV *x*-vector proposed in [19] uses eight hidden layers. Owing to the small dataset, we propose a similar but shallow architecture with just four hidden layers.

The modified x-vector architecture for spoof detection is shown in Figure 1. The first two layers are frame level layers and use TDNNs. These layers convert the input feature vectors into high-dimensional vectors by preserving temporal information. The third layer averages information across time by estimating mean and standard deviation, thereby converting the inputs of variable length into a fixed length, high-dimensional vector. The fourth hidden layer reduces this high-dimensional vector to a low-dimensional representation. The fifth (final) layer uses softmax activation and has only two output nodes. The network is trained discriminatively using binary labels as opposed to speaker labels in the case of *x*-vectors for ASV. After training *x*-vectors, ASV uses the low dimensional embeddings to verify the speaker. Since spoof detection is a binary classification problem, the network posteriors are used to determine the final score.

Instead of training the network using standard cross-entropy error, the focal loss function is used in this work. Focal loss was first proposed for object detection task in [12]. The focal loss reshapes the cross-entropy loss such that it gives more importance for hard-to-classify and misclassified exam-
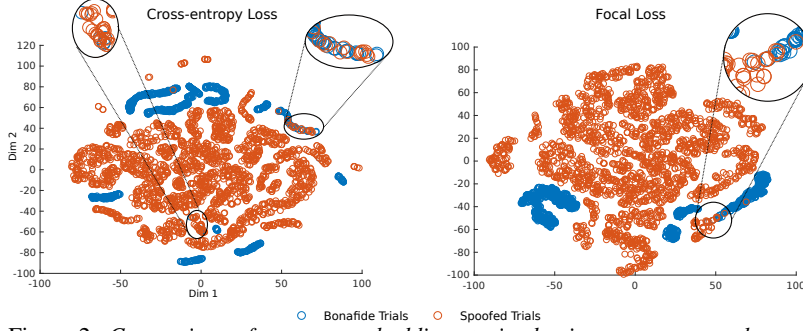
Figure 2: *Comparison of x-vector embeddings trained using cross-entropy loss and focal loss. The LA development subset of ASV-spoof-2019 dataset is used to generate this plot.*

Table 2: *List of developed systems.*

Type	System	System Name	
		GMM	x-Vector
Single	Baseline	G-CQCC	x-CQCC
		G-LFCC	x-LFCC
		G-IMFCC	x-IMFCC
		G-LFBE	x-LFBE
DLFS	Primary	G-Prim	x-Prim
	Contrastive-1	G-C1	x-C1
	Contrastive-2	G-C2	x-C2

ples. Hence, the focal loss is a better loss function for the class imbalance problem. The focal loss is estimated as shown in Equation 1.

$$\mathcal{F}(p, y) = -\alpha \left[y \, (1-p)^\gamma \, log(p) \, + \, (1-y) \, (p)^\gamma \, log(p) \right] \tag{1}$$

In Equation 1, y is the ground truth class label, and p is the posterior probability given by a neural network. α and γ are hyper-parameters in this loss function. Setting γ to zero reduces focal loss to the standard cross-entropy loss. In Figure 2, the 2D representation of DNN embeddings obtained from x-vectors trained using the cross-entropy loss and the focal loss are compared. DNN embeddings are converted to 2D space using t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm [20]. It can be observed that focal loss produces better embeddings with lesser inter-class overlap than the standard cross-entropy error.

In ASV, the *x*-vector architecture uses the raw filter bank energies as the input. Borrowing from the ASV approaches, the same filter bank energies were given as input to the *x*-vector framework. As the performance was poor, the focus is shifted to use different features for building a better classifier.

6. Spoof detection systems (SDS)

Several attempts have been made to train an efficient classifier for spoof detection. The most common classifiers used for the purpose are GMMs and DNNs. Although there are few works with SVMs [21] and i-vectors [22], the performance is worse than that of the GMM and DNN classifiers. Hence in this work, we use both GMM and DNN classifiers to detect spoofed trials. GMM-based systems with a set of features were explored, and best performing four systems were submitted to the ASV-spoof-2019 challenge. The *x*-vector based SDS were developed post-challenge. The performance of the *x*-vector systems is compared with the submitted GMM-based SDS using development data.

6.1. Single feature systems

GMM classifier has been the baseline system for all the ASV spoof challenges conducted from 2015 to 2019. Bonafide and spoofed trials from the training subset are used to train two GMMs, one for the bonafide (λ_B) and other for the spoofed class (λ_S). During testing, a trial t, is given to λ_B and λ_S, and

the log-likelihood (Λ) difference is computed as

$$\mathcal{S}(t) = \Lambda(\lambda_B(t)) - \Lambda(\lambda_S(t)) \tag{2}$$

The log-likelihood difference is considered as the final score for the trial t. This simple classifier gave an EER of 1.44% and 7.82% on the evaluation data of ASV-spoof-2015 [5] and ASV-spoof-2017 [8] respectively. GMM SDS with a set of cepstral coefficients and filterbank energies were explored for the ASV-spoof-2019 challenge. The GMM systems with constant-Q cepstral coefficients (CQCC) [23], inverse Mel frequency cepstral coefficients (IMFCC) [24], linear frequency cepstral coefficients (LFCC) [25], and linear filterbank energy (LFBE) gave better performance than few other features like Mel frequency cepstral coefficients (MFCC), inverse Mel filterbank energies (IMFBE), and Mel filterbank energies (MFBE). To compare the performance of *x*-vector systems with that of the baseline GMM systems, *x*-vector systems were also developed with the same set of features.

6.2. Feature switching systems

Almost every spoof detection system uses a score fusion of many single feature based system as the primary system [4, 6]. This clearly shows that different features are required to detect different spoofing conditions. Instead of the conventional score fusion approach, a decision-level feature switching (DLFS) approach proposed in [8] is used here. For a given trial, DLFS essentially chooses the decision score from a set of individual features, that has maximum discrimination between the bonafide and the spoofed model. In this work, DLFS is implemented with four best performing individual feature based system for both GMM and *x*-vector frameworks. The list of systems developed for this work is listed in Table 2. Features used in primary and contrastive DLFS systems vary for logical access and physical access.

7. Result Analysis

The *x*-vector neural network for LA and PA spoof detection is trained only on the corresponding training subsets. To avoid the problem of over-fitting, twenty percentage of training data is used as the validation subset. Since the ground truth of the evaluation subset is not released, the performance of the *x*-vector systems is analyzed only with the development subset. The performance of all spoof detection systems on the development and evaluation data are listed in Table 3. The best performing system is based on the min-t-DCF metric [9, 10]. The performance

Table 3: *Performance of various spoof detection systems*

| System Type | System Name | Feature | Logical access | | | | | Physical access | | | | | Feature |
| | | | Development Data | | | Evaluation Data | | Development Data | | | Evaluation Data | | |
			min-t-DCF	EER	Acc	min-t-DCF	EER	min-t-DCF	EER	Acc	min-t-DCF	EER	
Single	G-CQCC	CQCC	**0.0123**	**0.43**	97.02	0.2366	9.57	**0.1953**	**9.87**	**89.6**	0.2454	11.04	CQCC
	x-CQCC		0.0164	0.54	**99.69**	-	-	0.3039	12.98	85.50	-	-	
	G-LFCC	LFCC	0.0663	2.71	83.71	0.2116	8.09	0.2555	11.96	90.55	0.3017	13.54	LFCC
	x-LFCC		**0.0062**	**0.28**	**99.72**	-	-	**0.1231**	**4.53**	**94.66**	-	-	
	G-IMFCC†	IMFCC	**0.0012**	**0.04**	95.7	-	-	0.2078	9.19	91.41	0.3085	12.10	IMFCC
	x-IMFCC		0.0285	1.08	**99.40**	-	-	**0.1396**	**5.28**	**94.60**	-	-	
	G-LFBE*	LFBE	**0.0077**	**0.32**	**98.9**	0.2059	10.65	0.2581	11.47	85.99	-	-	LFBE
	x-LFBE		0.0561	1.88	98.82	-	-	**0.1818**	**7.39**	**92.07**	-	-	
DLFS	G-Prim*†	CQCC | LFBE	**0.0002**	**0.01**	99.94	0.1333	6.14	0.1888	8.17	91.00	0.2767	11.28	CQCC | IMFCC | LFBE
	x-Prim		0.0139	0.47	99.85	-	-	**0.1236**	**4.85**	92.28	-	-	
	G-C1*†	IMFCC | LFBE	**0.0003**	**0.04**	99.95	0.1565	6.46	0.1972	7.53	**92.34**	0.2309	9.33	CQCC | LFCC
	x-C1		0.0040	0.16	**99.96**	-	-	**0.1226**	**4.56**	92.67	-	-	
	G-C2*†	CQCC | LFCC	**0.0013**	**0.04**	99.94	0.2139	9.04	0.2329	8.48	**91.52**	0.3058	11.34	CQCC | LFBE
	x-C2		0.0142	0.47	99.88	-	-	**0.1821**	**7.54**	89.83	-	-	
	G-DLFS	CQCC| IMFCC | LFCC	**0.0026**	0.19	98.29	-	-	0.1548	7.61	91.80	-	-	CQCC | IMFCC | LFCC
	x-DLFS		0.0033	**0.14**	**99.92**	-	-	**0.1171**	**4.13**	**93.78**	-	-	

Systems marked with * and † were submitted to ASV-spoof-2019 challenge under LA and PA conditions respectively. The symbol '|' represenCQCClus. Scts exoreive Os from feature A (OR) B will be chosen for each trial.

of GMM systems submitted to the challenge is reported on both development and evaluation subsets. G-IMFCC and G-LFBE are the single systems submitted for PA and LA respectively, whereas G-CQCC and G-LFCC are the single feature based baseline systems provided along with the challenge dataset.

Results reported in Table 3 shows that all the GMM based SDS under LA category performs quite well compared to that of the PA category. From the performance in terms of min-t-DCF and EER, we can infer that GMM is a better classifier, for the detection of LA-based spoof attacks than *x*-vector in most of the cases.

Out of all the development trials of LA, the best *x*-vector system (x-C1) and the corresponding GMM system (G-C1), have only 10 and 11 misclassifications respectively. Although there is considerable variation in terms of min-t-DCF and EER, the number of misclassified trials are almost same in both the x-C1 and G-C1. Hence, with the results on development subset, neither GMM nor *x*-vector can be chosen as the better classifier for LA based spoof detection. In terms of min-t-DCF, G-Prime system gave a relative improvement of 98.37% over the best baseline (G-CQCC) provided for ASV-spoof-2019 challenge.

For spoof detection under PA category, the GMM spoof detection systems using different features did not surpass the best baseline (G-CQCC) system. Unlike LA category, SDS with *x*-vector outperforms all the GMM based SDS. One possible reason could be that, unlike LA, the PA category have enough amount of data (refer Table 1) to train the neural network. On comparing single feature based systems, the best performing *x*-vector system (x-LFCC) surpasses best GMM system (G-CQCC) with a relative improvement of 36.97% in terms of min-t-DCF. Since *x*-vector performs well for all the cases, we can conclude it as a more suitable classifier for detecting physical access spoofing. The performance of the SDS is further improved by applying DLFS as shown in the Table 3.

Apart from the systems submitted to the challenge, DLFS with a new feature combination is reported in the table as G-DLFS and x-DLFS. On comparing with the best baseline system (G-CQCC), this new x-DLFS system give best performance for PA with a relative improvement of 40.04% on min-t-DCF.

From the result analysis of both LA and PA, we can con-

clude that *x*-vector framework can be a potential model to detect all types of spoofing attacks. It also justifies our assumption that *x*-vector better identifies the traces of spoof mechanism in the spoofed utterances than the GMM. Moreover, since *x*-vector is the current state-of-the-art for ASV, *x*-vector based spoof detection, will help us to make a common NN framework for spoof detection as well as speaker recognition.

8. Conclusion

Spoofed utterances contain traces of approaches used to generate them. The ability of *x*-vector based NN to capture the utterance level information is established in the field of speaker verification. Hence, in this work, an attempt has been made to develop spoof detection systems using *x*-vector framework. A shallow NN with focal-loss as the loss function is proposed as the *x*-vector architecture for spoof detection. On ASV-spoof-2019 dataset, the proposed *x*-vector based SDS outperforms all the GMM based SDS in case of PA, whereas GMM based SDS performs well for LA in most of the cases. Further, DLFS paradigm is used to improve the performance of single feature based SDS. The best performing *x*-vector SDS with DLFS outperforms the best performing GMM-DLFS SDS with a relative improvement of 34.53% for PA in terms of min-t-DCF respectively.

9. Acknowledgements

We would like to thank the ASV-Spoof-2019 organizers for providing the new dataset for the spoof detection task.

10. References

[1] S. J. Elliott, *Zero Effort Forgery*. Boston, MA: Springer US, 2009, pp. 1411–1414.

[2] Z. Wu *et al.*, "Spoofing and countermeasures for speaker verification: A survey," *Speech Communication*, vol. 66, pp. 130 – 153, 2015.

[3] T. B. Patel and H. A. Patil, "Significance of Source-Filter Interaction for Classification of Natural vs. Spoofed Speech," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 644–659, June 2017.

[4] Z. Wu *et al.*, "ASVspoof: The Automatic Speaker Verification Spoofing and Countermeasures Challenge," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 588–604, June 2017.

[5] T. B. Patel and H. A. Patil, "Cochlear Filter and Instantaneous Frequency Based Features for Spoofed Speech Detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 618–631, June 2017.

[6] T. Kinnunen *et al.*, "The ASVspoof 2017 challenge: Assessing the limits of replay spoofing attack detection," in *INTERSPEECH*, Aug 2017, pp. 1–6.

[7] F. Tom, M. Jain, and P. Dey, "End-to-end audio replay attack detection using deep convolutional networks with attention," in *INTERSPEECH*, 2018, pp. 681–685. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2018-2279

[8] Saranya M. S. and Hema A. Murthy, "Decision-level feature switching as a paradigm for replay attack detection," in *INTERSPEECH*, 2018, pp. 686–690.

[9] T. Kinnunen, K. A. Lee *et al.*, "t-DCF: a detection cost function for the tandem assessment of spoofing countermeasures and automatic speaker verification," in *Odyssey, The Speaker and Language Recognition Workshop*, June 2018.

[10] T. Kinnunen, K.-A. Lee, H. Delgado, N. W. D. Evans, M. Todisco, M. Sahidullah, J. Yamagishi, and D. A. Reynolds, "t-dcf: a detection cost function for the tandem assessment of spoofing countermeasures and automatic speaker verification," *CoRR*, vol. abs/1804.09618, 2018.

[11] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-Vectors: Robust DNN Embeddings for Speaker Recognition," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5329–5333, 2018.

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[13] Zhizheng Wu and Tomi Kinnunen and Nicholas Evans and Junichi Yamagishi , "Automatic Speaker Verification Spoofing and Countermeasures Challenge (ASVspoof)," Feb 2015. [Online]. Available: http://www.spoofingchallenge.org/index2015.html

[14] G. Lavrentyeva *et al.*, "Audio replay attack detection with deep learning frameworks," in *INTERSPEECH*, Aug 2017, pp. 82–86.

[15] O. Russakovsky, J. Deng *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[16] T. Kinnunen *et al.*, "Automatic Speaker Verification Spoofing and Countermeasures Challenge (ASVspoof)," Feb 2017. [Online]. Available: http://www.spoofingchallenge.org/index2017.html

[17] ASVspoof consortium, "ASVspoof 2019:Automatic Speaker Verification Spoofing andCountermeasures Challenge Evaluation Plan," Jan 2019. [Online]. Available: http://www.asvspoof.org/asvspoof2019/asvspoof2019_evaluation_plan.pdf

[18] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, May 2011.

[19] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, "Deep neural network embeddings for text-independent speaker verification." in *Interspeech*, 2017, pp. 999–1003.

[20] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008. [Online]. Available: http://www.jmlr.org/papers/v9/vandermaaten08a.html

[21] S. Novoselov, A. Kozlov, G. Lavrentyeva, K. Simonchik, and V. Shchemelinin, "Stc anti-spoofing systems for the asvspoof 2015 challenge," in *ICASSP*, March 2016, pp. 5475–5479.

[22] E. Khoury, T. Kinnunen, A. Sizov, Z. Wu, and S. Marcel, "Introducing i-vectors for joint anti-spoofing and speaker verification," in *INTERSPEECH*, 2014, pp. 61–65.

[23] M. Todisco *et al.*, "A New Feature for Automatic Speaker Verification Anti-Spoofing: Constant Q Cepstral Coefficients," in *The Speaker and Language Recognition Workshop, ODYSSEY*, June 2016.

[24] S. Chakroborty, A. Roy, and G. Saha, "Improved closed set text-independent speaker identification by combining mfcc with evidence from flipped filter banks," *International Journal of Signal Processing*, vol. 4, no. 2, pp. 114–122, 2007.

[25] X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson, and S. Shamma, "Linear versus Mel frequency cepstral coefficients for speaker recognition," in *2011 IEEE Workshop on Automatic Speech Recognition Understanding*, Dec 2011, pp. 559–564.