CPAUG: REFINING COPY-PASTE AUGMENTATION FOR SPEECH ANTI-SPOOFING

Linjuan Zhang[1,2], Kong Aik Lee[3,2*], Lin Zhang[4], Longbiao Wang[5], Baoning Niu[1,⋆]

[1]*Taiyuan University of Technology, Taiyuan, China*

[2]*Institute for Infocomm Research, A*STAR, Singapore*

[3]*The Hong Kong Polytechnic University, Hong Kong, China*

[4]*National Institute of Informatics, Tokyo, Japan*

[5]*Tianjin University, Tianjin, China*

Abstract

Innovation

The proposed **CpAug** method represents a refined augmentation specifically for anti-spoofing.

① Unlike the conventional copy-paste that only concatenates utterances, CpAug also substitutes segments.

② Explore four blending strategies that consider the same / different speakers and spoofing attack types when creating augmented instances.

③ CpAug integrates basic signal processing to mitigate problems from directly concatenating signals.

④ Compare the performance of the proposed CpAug with now widely-used Rawboost, highlighting their synergistic benefits.

Results

① Compared with no data augmentation, the **CpAug** with *substitution* policy leads to relative improvements of **43%** and **38%** on the ASVspoof' 19LA and 21LA, respectively.

② The **CpAug** and Rawboost synergize effectively, achieving an EER of **2.91%** on ASVspoof' 21LA.

CpAug: Refing Copy-paste Augmentation

Methodology

- **The Conventional Copy-paste Augmentation**

$$X_{\mathrm{CP}} = f_{\mathrm{CP}}\left(X_1, X_2\right) = X_1 \oplus X_2$$

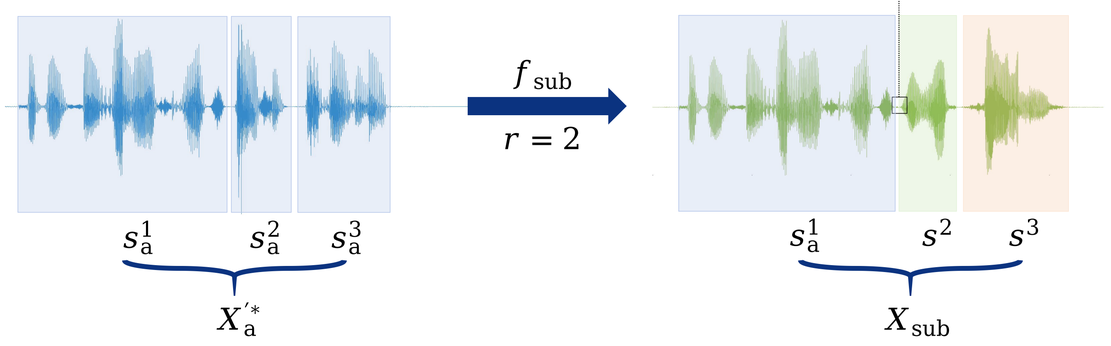
- **Substitution Policy**

$$X_{\mathrm{sub}} = f_{\mathrm{sub}}\left(X_{\mathrm{a}}'^{*}, \mathscr{S}, r\right)$$

$$= \left\{ s_{\mathrm{sub}}^{i} = \begin{cases} s^{j} & j \in R \quad \text{if} \quad i \in Q, \\ s_{\mathrm{a}}^{i} & \text{otherwise.} \end{cases} \right\}_{i=1}^{C}$$

$$\mathscr{S} \triangleq \left\{ s^{i} : i = 1, \ldots, M \right\}$$

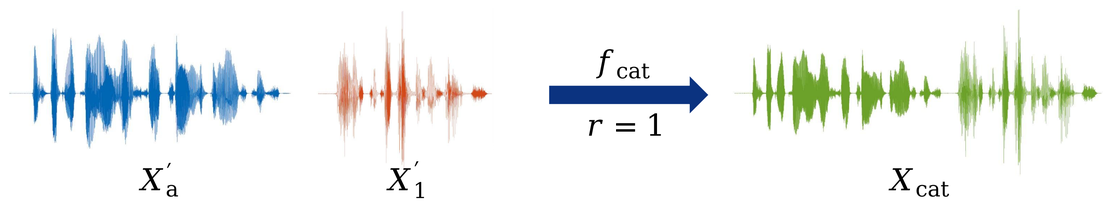
(cross-correlation, overlap-add)



- **Concatenation Policy**

$$X_{\mathrm{cat}} = f_{\mathrm{cat}}\left(X_{\mathrm{a}}', \mathscr{D}, r\right) = X_{\mathrm{a}}' \oplus X_1' \oplus X_2' \oplus \cdots \oplus X_r'$$

$$\mathscr{D} \triangleq \left\{ X_i' : i = 1, \ldots, K \right\}$$



Different Blending Strategies

Tab. 1. Details of the augmented datasets with four blending strategies.

	Names	Blending Strategies		Ave_ Dur (s)	Ratio (%)
		Attack	Speaker		
Concatenation	cat-satt-sspk	same	same	6.62	52.14
	cat-datt-sspk	different	same	6.67	51.91
	cat-satt-dspk	same	different	6.62	52.14
	cat-datt-dspk	different	different	6.66	51.83
Substitution	sub-satt-sspk	same	same	3.23	45.88
	sub-datt-sspk	different	same	3.25	45.20
	sub-satt-dspk	same	different	3.23	45.78
	sub-datt-dspk	different	different	3.23	45.06

Experimental Results

Tab. 2. Performance in EER (%) and t-DCF on 19LA test set with different policies.

DA	r	Back-end	EER	t-DCF
No	0		0.93	0.0285
conve CP	1		1.22	0.0410
cat-satt-dspk	1	AASIST	**0.86**	**0.0253**
sub-satt-dspk*	1		0.90	0.0292
sub-satt-dspk	[1,16]		0.88	0.0270
No	0		1.35	0.0425
conve CP	1		1.40	0.0440
cat-satt-dspk	1	Rawformer	1.09	0.0341
sub-satt-dspk*	1		1.02	0.0332
sub-satt-dspk	[1,16]		**0.77**	**0.0253**

Tab. 3. EER (%) and t-DCF on 19LA and 21LA tests with different blending strategies using Rawformer.

DA	r	19LA		21LA	
		EER	t-DCF	EER	t-DCF
No	-	1.35	0.0425	6.38	0.3328
cat-satt-sspk	1	1.16	0.0355	**4.41**	**0.3019**
cat-datt-sspk	1	1.22	0.0386	5.22	0.3309
cat-satt-dspk	1	**1.09**	**0.0341**	4.85	0.3086
cat-datt-dspk	1	1.28	0.0431	5.76	0.3380
sub-satt-sspk	[1,16]	0.88	0.0288	**3.93**	**0.2851**
sub-datt-sspk	[1,16]	1.02	0.0336	4.23	0.2909
sub-satt-dspk	[1,16]	**0.77**	**0.0253**	4.36	0.2993
sub-datt-dspk	[1,16]	1.12	0.0371	5.13	0.3086

- **conve CP:** the conventional copy-paste augmentation
- ***r*:** the number of concatenated utterances or substituted segments

- sub-satt-sspk (Rawboost): Rawboost augmentation before 'sub-satt-sspk'
- Rawboost (sub-satt-sspk): Rawboost augmentation after 'sub-satt-sspk'
- Rawboost,sub-satt-sspk: independent augmentation using Rawboost and *substitution* strategies, separately

Tab. 4. Performance of combining the CpAug and Rawboost on the 21LA test set.

DA	Back-end	EER	t-DCF
No	AASIST	10.51	0.4884
Rawboost		7.60	0.2601
sub-satt-sspk		7.31	0.2488
sub-satt-sspk(Rawboost)		8.59	0.2853
Rawboost(sub-satt-sspk)		**5.56**	**0.1776**
Rawboost,sub-satt-sspk		5.60	0.1850
No	Rawformer	6.38	0.3328
Rawboost		4.02	0.2918
sub-satt-sspk		3.93	0.2851
sub-satt-sspk(Rawboost)		3.98	0.2740
Rawboost(sub-satt-sspk)		**2.91**	**0.2617**
Rawboost,sub-satt-sspk		3.61	0.2825

https://github.com/zlin0/CpAug

zhanglinjuan@tyut.edu.cn

Conclusions

✓ We proposed the **CpAug** tailored for anti-spoofing using *concatenation* and *substitution* policies and found it performs well on the ASVspoof' 19LA and 21LA tests.

✓ We explored four different blending strategies and found that using the same spoofing attack type achieves the best performance.

✓ The proposed **CpAug** and Rawboost work well together, with the Rawboost(sub-satt-sspk) giving the best results.

Acknowledgements

We would like to thank Prof. Junichi Yamagishi for his invaluable suggestions.

This work was supported by the National Natural Science Foundation of China (Grant No. 62072326), the Key Research and Development Plan of Shanxi Province (202203021212282), in part by the China Scholarship Council (No. 202206930011), and part by Japan Science and Technology Agency Grants JPMJFS2136.

The work was done during Linjuan Zhang's internship at A*STAR.