From Sharpness to Better Generalization for Speech Deepfake Detection

Wen Huang[1], Xuechen Liu[2], Xin Wang[2], Junichi Yamagishi[2], Yanmin Qian[1]

[1]AudioCC Lab, Shanghai Jiao Tong University, China

[2]National Institute of Informatics, Japan

INTERSPEECH 2025
17-21 AUGUST - ROTTERDAM

Motivation

- **Problem**: Models for ***Speech Deepfake Detection (SDD)*** often fail to generalize across unseen domains.
- **Gap**: Lack of a theoretical framework to explain or predict ***generalization***.
- **Proposal**: Use ***sharpness*** as a proxy to understand and improve generalization.

Key Questions

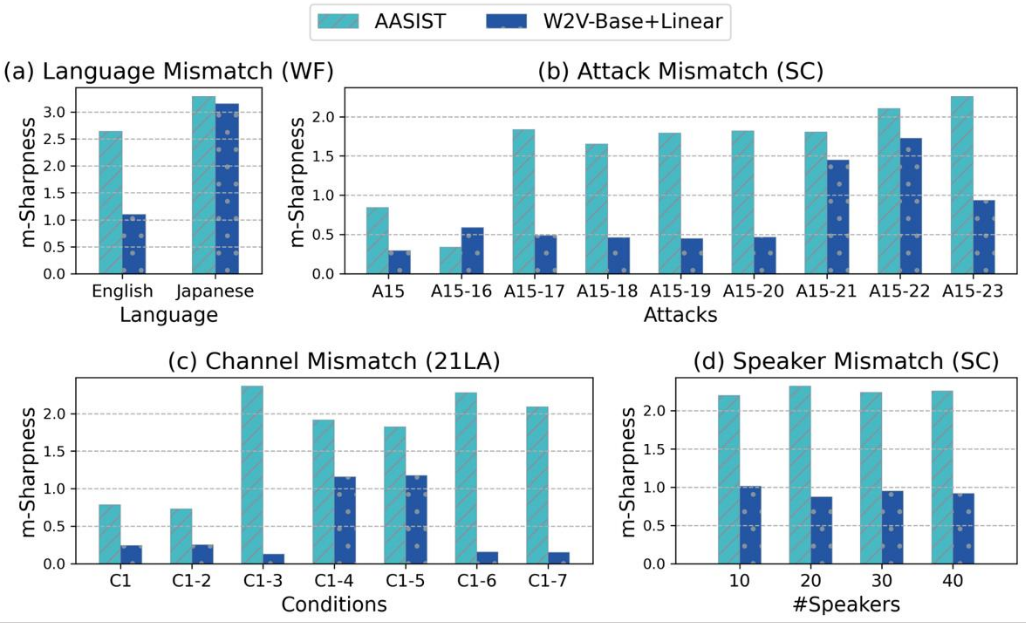
- Can ***sharpness*** serve as a ***theoretical indicator*** for generalization in SDD?
- Can ***Sharpness-Aware Minimization (SAM)**** enhance generalization performance across diverse datasets?

Sharpness & Domain Mismatch

- **Sharpness**: Measures model sensitivity to parameter perturbations.

$$s(w, S) \triangleq \max_{\|\epsilon\|_2 \leq \rho} \frac{1}{|S|} \sum_{i:(x_i, y_i) \in S} (\ell_i(w+\epsilon) - \ell_i(w))$$

- Sharpness increases under ***unseen conditions***: languages, spoofing attacks, channel effects, (but not speaker variability)
- Intuitively, ***lower sharpness*** -> ***less sensitivity*** to data -> ***better generalization***



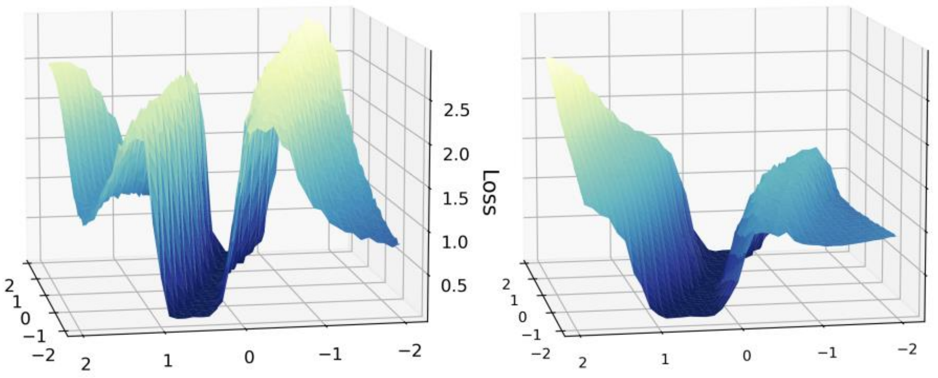
Sharpness-Aware Minimization

- **Objective**: *original loss + regularization*

$$L_S(w) + \lambda\|w\|_2^2 + \underbrace{[\max_{\|\epsilon\|_2 \leq \rho} L_S(w+\epsilon) - L_S(w)]}_{\textit{sharpness-aware component}}$$

- Simultaneously minimize the loss value and its sharpness, achieving flatter loss landscapes.

Loss Landscape of model trained with Adam (L) or SAM (R)



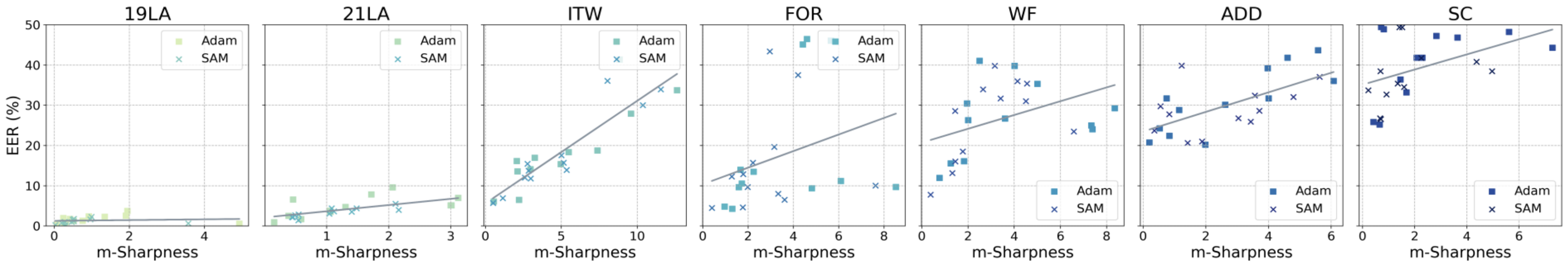
Experiments and Analysis

- **Experimental Setup: *Train set*:** ASVspoof 2019 LA; ***Test set***: 8 datasets; ***Models***: AASIST | W2V+Linear or AASIST; ***Training:*** Cross-entropy loss, RawBoost augmentation, Adam or SAM optimizer

Equal Error Rate (EER) % (↓) of different models trained with Adam or SAM

Model	Optimizer	19LA	21LA	21DF	ITW	FOR	WF	ADD	SC
AASIST	Adam	1.96 $\pm$ 0.61	8.10 $\pm$ 1.33	21.92 $\pm$ 2.61	34.31 $\pm$ 6.73	45.85 $\pm$ **0.68**	38.68 $\pm$ 3.02	40.48 $\pm$ 3.98	45.27 $\pm$ **3.01**
	SAM	**1.71 $\pm$ 0.55**	**4.62 $\pm$ 0.81**	**19.58 $\pm$ 1.18**	**33.34 $\pm$ 3.07**	**33.51** $\pm$ 12.37	**33.85 $\pm$ 2.15**	**33.84 $\pm$ 2.80**	**43.98** $\pm$ 4.71
W2V-Base+Linear	Adam	1.78 $\pm$ 0.51	5.82 $\pm$ 1.01	13.10 $\pm$ 3.00	16.85 $\pm$ 2.74	12.01 $\pm$ **2.07**	**30.34 $\pm$ 6.71**	30.47 $\pm$ 9.03	34.98 $\pm$ **1.66**
	SAM	**1.21 $\pm$ 0.34**	**3.39 $\pm$ 0.89**	**11.93 $\pm$ 0.55**	**13.66 $\pm$ 1.82**	**9.57** $\pm$ 2.78	36.95 $\pm$ 7.40	**24.29 $\pm$ 3.57**	**34.87** $\pm$ 2.97
W2V-Base+AASIST	Adam	2.81 $\pm$ 0.79	4.78 $\pm$ 0.57	10.37 $\pm$ 1.19	18.29 $\pm$ **0.47**	11.11 $\pm$ 1.73	**27.79** $\pm$ 3.29	36.22 $\pm$ 5.28	45.89 $\pm$ 3.56
	SAM	**1.32 $\pm$ 0.38**	**3.12 $\pm$ 0.39**	**10.00 $\pm$ 0.76**	**16.21** $\pm$ 2.02	**7.92 $\pm$ 1.37**	34.29 $\pm$ **2.88**	**28.48 $\pm$ 3.73**	**34.83 $\pm$ 0.46**
W2V-Large+Linear	Adam	1.37 $\pm$ 0.40	3.11 $\pm$ 0.64	6.79 $\pm$ 0.38	14.96 $\pm$ 1.41	**13.24** $\pm$ 2.44	23.97 $\pm$ 7.30	30.97 $\pm$ 9.91	48.35 $\pm$ **1.40**
	SAM	**0.88 $\pm$ 0.10**	**2.58 $\pm$ 0.37**	**6.37 $\pm$ 0.30**	**12.22 $\pm$ 1.41**	14.65 $\pm$ **1.79**	**16.69 $\pm$ 1.59**	**37.31 $\pm$ 6.68**	**42.89** $\pm$ 5.73
W2V-Large+AASIST	Adam	1.22 $\pm$ 0.60	4.53 $\pm$ 0.99	7.17 $\pm$ **0.17**	16.36 $\pm$ **1.80**	11.58 $\pm$ 1.81	27.89 $\pm$ 2.76	33.60 $\pm$ 3.40	**39.62 $\pm$ 4.50**
	SAM	**1.04 $\pm$ 0.47**	**3.60 $\pm$ 0.16**	**6.82** $\pm$ 0.32	**15.46** $\pm$ 2.09	**10.02 $\pm$ 0.82**	**25.12 $\pm$ 1.47**	**31.41 $\pm$ 3.28**	39.67 $\pm$ 5.14
W2V-XLSR+Linear	Adam	0.34 $\pm$ 0.06	**1.32 $\pm$ 0.35**	4.27 $\pm$ **0.43**	6.00 $\pm$ **0.51**	4.55 $\pm$ 1.19	9.87 $\pm$ 3.24	22.85 $\pm$ 2.78	25.82 $\pm$ **1.89**
	SAM	**0.20 $\pm$ 0.05**	1.87 $\pm$ 0.39	**3.38** $\pm$ 0.47	**5.99** $\pm$ 0.80	**3.69 $\pm$ 0.90**	**7.66 $\pm$ 1.24**	**21.71 $\pm$ 2.08**	**25.65** $\pm$ 2.74
W2V-XLSR+AASIST	Adam	0.34 $\pm$ 0.13	1.85 $\pm$ 0.25	3.61 $\pm$ **0.32**	6.89 $\pm$ 1.19	**4.56 $\pm$ 0.72**	16.92 $\pm$ 7.36	**19.67** $\pm$ 1.67	**27.50 $\pm$ 1.98**
	SAM	**0.25 $\pm$ 0.12**	**1.71 $\pm$ 0.27**	**3.44** $\pm$ 0.54	**6.34 $\pm$ 0.62**	5.18 $\pm$ 1.48	**14.36 $\pm$ 4.74**	21.36 $\pm$ **0.59**	29.93 $\pm$ 3.30

- **Generalization with SAM: *Consistent EER reduction*** across most models & datasets; ***Most gains*** seen in ***mismatched*** conditions (21LA, ITW, ADD, WF); ***SSL + SAM*** outperforms all other combinations.



- **Sharpness ↔ Generalization (Correlation Analysis)**:
- Strongest in mismatched domains: ITW, 21LA, ADD
- Moderate in FOR, WF, SC; Weakest in in-domain (19LA)

Metric	19LA	21LA	ITW	FOR	WF	ADD	SC
PCC	0.13	0.65	0.89	0.30	0.40	0.65	0.43
SRCC	0.53	0.77	0.87	0.40	0.52	0.62	0.49
KTAU	0.45	0.63	0.72	0.32	0.36	0.34	0.34

Conclusion

- Sharpness increases under domain shifts, correlates with performance → useful indicator of generalization.
- SAM reduces sharpness → more robust models and better generalization.

Paper

Code