HỘI THẢO KHOA HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG NĂM 2025
THE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY

Research Article:

Enhancing Vietnamese Audio Anti-Spoofing with the AASIST Model and Data Augmentation

Ngoc-Thao Tran
The Faculty of Engineering and Technology, Baclieu University

Duy-Quy Thai
The Faculty of Information Technology, Dalat University

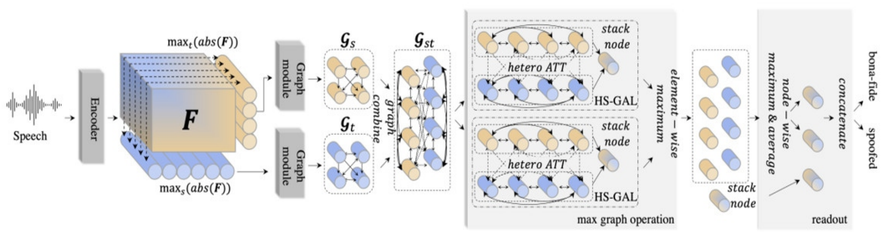
Problem Statement

- Speech deepfakes and spoofing attacks threaten voice authentication systems;
- Vietnamese language is underrepresented in current anti-spoofing solutions;
- Existing models trained on English perform poorly on Vietnamese.

Proposed Solution

- Fine-tune AASIST model on Vietnamese speech;
- Apply advanced data augmentation techniques (e.g., RawBoost, noise injection, MP3 compression).

Proposed Solution pipeline



Fine-tune on Vietnamese speech

Retraining Strategy
- Freeze lower layers of the network;
- Fine-tune upper layers on Vietnamese data.

Data Augmentation Techniques
- RawBoost approach;
- Mp3 compression;
- Additive white noise;
- Loudness normalization.

AASIST model:
- Combines Graph Attention with Spectro-Temporal features;
- Processes raw audio directly;
- Learns both spectral and temporal dependencies.

Dataset

Vietnamese Spoofing-aware Speaker Verification (VSASV):
- 340,000 utterances;
- 1,400 speakers;
- Diverse accents, male/female balance;
- Real and spoofed audio (replay, voice conversion, adversarial attacks).

Metrics evaluate

- **Equal Error Rate (EER)**: Measures the trade-off between false acceptance and rejection;
- **Accuracy and F1-score**: Indicate classification performance;
- **t-DCF** Assesses the impact when integrated with an ASV system.

Conclusion

- AASIST effectively adapts to Vietnamese with fine-tuning;
- Data augmentation enhances real-world robustness;
- Strong potential for secure Vietnamese voice authentication systems.

Results

Performance comparison in EER

System	Dev. set	Test set
AASIST-L	49.56	50.45
AASIST	49.92	50.79
AASIST fine tune without augmentation	11.10	11.12
AASIST fine tune augmentation	5.54	5.28

Performance comparison on test set

System	Acc	min t-DCF	F1
AASIST-L	0.50	0.49	0.00
AASIST	0.5	0.49	0.00
AASIST fine tune without augmentation	0.88	0.10	0.88
AASIST fine tune augmentation	0.95	0.05	0.95

The results evaluation on the VSASV dataset

System	EER
ECAPA-TDNN (a)	22.03
ECAPA-TDNN (b)	15.58
ECAPA-TDNN & AASIST ©	8.27
AASIST fine tune without augmentation	10.85
AASIST fine tune augmentation	5.71

AASIST - Audio Anti-Spoofing using Integrated Spectro-Temporal Graph Attention