KAIST EE

2024.04.17 (Wed)

NOISY-ARCMIX

ADDITIVE NOISY ANGULAR MARGIN LOSS COMBINED WITH MIXUP FOR ANOMALOUS SOUND DETECTION

ICASSP 2024 KOREA

Soonhyeon Choi, Jung-Woo Choi

School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST)

Highlights

- **Noisy-ArcMix**
  - Anomalous sound detection objectives to achieve compact intra-class distribution and improved discrimination for anomalous samples
- **TAgram**
  - Temporally attended feature that captures which time frames are important for ASD model

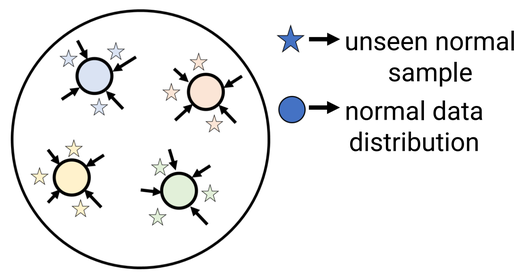
→ **Achieving state-of-the-art performance on DCASE 2020 Challenge Task 2 dataset**

Introduction

- **Anomalous Sound Detection (ASD)**
  - Identifying whether the sound from a target machine is normal or anomalous
  - Training only on normal data without anomalous data

- **Limitations of conventional models**

1) Failure to compact intra-class distribution due to less consideration for its surroundings when performing self-supervised classification task
2) Leading to misclassification when small perturbations are added to normal samples
3) Focusing only on local features without leveraging global temporal relations when extracting temporal information



- **Contributions**

Synthesizing new data around the distribution of normal data and applying inconsistent angular margins to secure representation space for anomalous samples

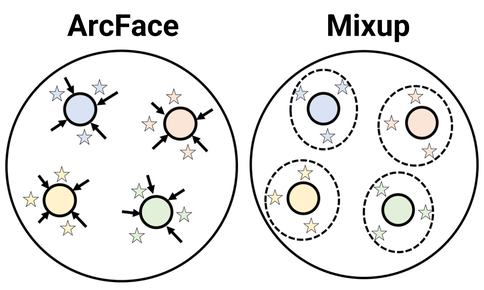
Proposed Methods

- **Background**

1) ArcFace loss (Additive angular margins)

$$\mathcal{L}_{\text{ArcFace}}(\mathbf{x}, \mathbf{y}) = -\mathbf{y}^\top \log \frac{e^{s\cos(\theta + m\mathbf{y})}}{\sum_{k=1}^{K} e^{s\cos(\theta_k + m y_k)}}$$

- Intra-class compactness through angular margins
- Lack of consideration for surroundings of each class

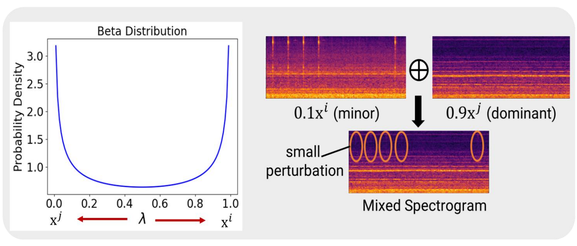


2) Mixup (Mixing of normal samples)

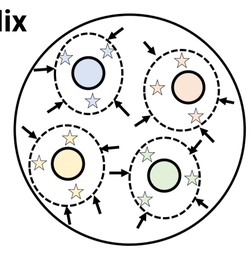
$$\mathbf{x}^{ij} = \lambda \mathbf{x}^i + (1-\lambda)\mathbf{x}^j$$

$$\mathcal{L}_{\text{CE}}(\mathbf{x}^{ij}, \mathbf{y}^{ij}) = -\mathbf{y}^{ij\top} \log \frac{e^{s\cos(\theta)}}{\sum_{k=1}^{K} e^{s\cos(\theta_k)}}$$

- Learning from surroundings of each class
- Unable to compact intra-class distribution



- **Noisy-ArcMix**



1) ArcMix loss (Combination of ArcFace and Mixup)

$$\mathcal{L}_{\text{AMix}}(\mathbf{x}^{ij}, \mathbf{y}^i, \mathbf{y}^j) = \lambda \mathcal{L}_{\text{AF}}(\mathbf{x}^{ij}, \mathbf{y}^i) + (1-\lambda)\mathcal{L}_{\text{AF}}(\mathbf{x}^{ij}, \mathbf{y}^j)$$

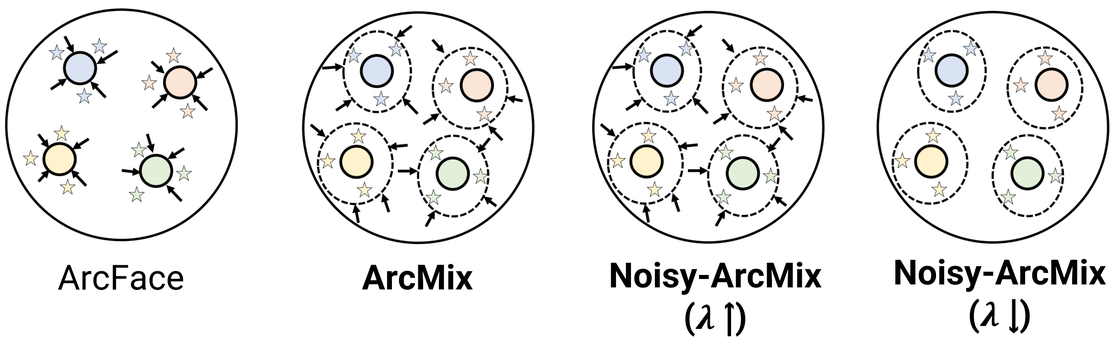
- Margin penalty added to both classes $\mathbf{y}^i$ and $\mathbf{y}^j$
- Enhanced intra-class compactness and expanding the representation space for anomalous data
- But leading to an intermingling effect between normal and anomalous samples due to compact intra-class distribution

2) Noisy-ArcMix loss (Inconsistent angular margins)

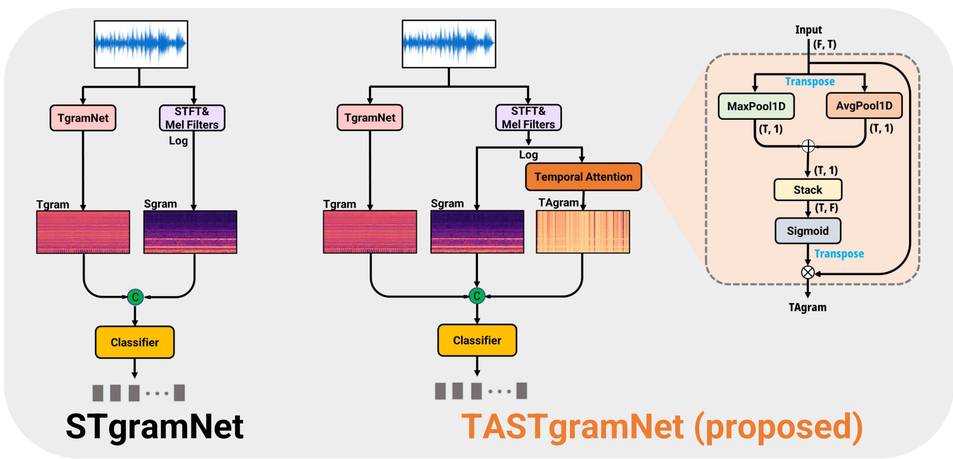
$$\mathcal{L}_{\text{NAMix}}(\mathbf{x}^{ij}, \mathbf{y}^{ij}) = -\mathbf{y}^{ij\top} \log \frac{e^{s\cos(\theta + m\mathbf{y}^i)}}{\sum_{k=1}^{K} e^{s\cos(\theta_k + m\mathbf{y}_k^i)}}$$

- Corrupted label, $\mathbf{y}^{ij} = \lambda \mathbf{y}^i + (1-\lambda)\mathbf{y}^j$
- Promoting the prediction of smaller probability $\lambda \mathbf{y}^i$ among $\mathbf{y}^{ij}$

- **Comparisons between ArcFace, ArcMix, and Noisy-ArcMix**



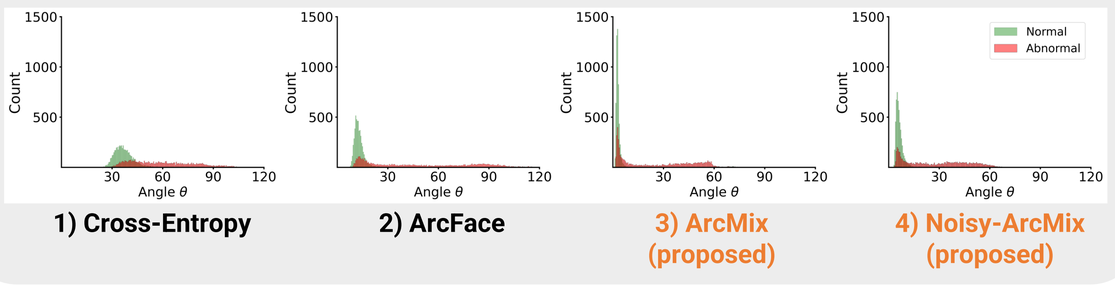
- **TASTgramNet Architecture**



- STgram: Sgram (log-Mel spectrogram) and Tgram (Temporal feature)
- Proposed: Temporally attended feature (TAgram)
- Capturing important temporal relations through pooling operations

Experimental Results

1) Angle distribution for each loss function



- Finding a better balance between inter-class compactness and discrimination for anomalous samples (Noisy-ArcMix)

2) Comparisons with baseline models

- Outperforming CLP-SCF and others, and achieving state-of-the-art results on DCASE 2020 Challenge Task 2 dataset

Methods	Fan		Pump		Slider		Valve		ToyCar		ToyConveyor		Average	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
IDNN	67.71	52.90	73.76	61.07	86.45	67.58	84.09	64.94	78.69	69.22	71.07	59.70	76.96	62.57
MobileNetV2	80.19	74.40	82.53	76.50	95.27	85.22	88.65	87.98	87.66	85.92	69.71	56.43	84.34	77.74
Glow-Aff	74.90	65.30	83.40	73.80	94.60	82.80	91.40	75.00	92.2	84.10	71.50	59.00	85.20	73.90
STgram (ArcFace)	94.04	88.97	91.94	81.75	99.55	97.61	99.64	98.44	94.44	87.68	74.57	63.6	92.36	86.34
CLP-SCF	96.98	93.23	94.97	**87.39**	**99.57**	**97.73**	99.89	99.51	95.85	**90.19**	75.21	62.79	93.75	88.48
TASTgram (NAMix)	**98.32**	**95.34**	**95.44**	85.99	99.53	97.50	**99.95**	**99.74**	**96.76**	90.11	**77.90**	**67.15**	**94.65**	**89.31**

3) Ablation studies on the effects of Noisy-ArcMix and TAgram

- Most of the performance improvement through Noisy-ArcMix
- Temporal attention helpful for improving ASD performance

Methods	Fan		Pump		Slider		Valve		ToyCar		ToyConveyor		Average	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
STgram (ArcFace)	94.04	88.97	91.94	81.75	99.55	97.61	99.64	98.44	94.44	87.68	74.57	63.60	92.36	86.34
TASTgram (ArcFace)	95.46	89.82	93.41	85.24	**99.78**	**98.81**	99.71	98.48	95.78	90.19	73.16	63.17	92.88	87.62
STgram (NAMix)	97.20	93.56	94.89	**87.07**	99.51	97.44	99.88	99.38	96.55	**90.41**	77.57	64.54	94.27	88.73
TASTgram (NAMix)	**98.32**	**95.34**	**95.44**	85.99	99.53	97.50	**99.95**	**99.74**	**96.76**	90.11	**77.90**	**67.15**	**94.65**	**89.31**

Conclusion

- We propose a training loss that gains generalization ability through the use of inconsistent angular margins.
- We propose temporally attended features to make a model focusing on global temporal relations.

KAIST csh5956@kaist.ac.kr, jwoo@kaist.ac.kr

IEEE Signal Processing Society