An ensemble approach to anomalous sound detection based on Conformer-based autoencoder and binary classifier incorporated with metric learning

Ibuki Kuroyanagi[1,2], Tomoki Hayashi[1,2], Yusuke Adachi[1,2], Takenori Yoshimura[1], Kazuya Takeda[2], Tomoki Toda[2]
[1]Human Dataware Lab. Co., Ltd. / [2]Nagoya University

INRODUCTION

✓ **The difficulty of anomalous sound detection (ASD)**

- Anomalous sounds are rarely occur
- The type of anomalous sounds are very diverse
- Real world environments are often changing and complicated

✓ **It is desirable**

- To train ASD models without using anomalous data
- To develop models that can detect anomalous sounds even when the normal state distribution is changed (i.e., after domain shift)

1. Propose completely different two ASD methods

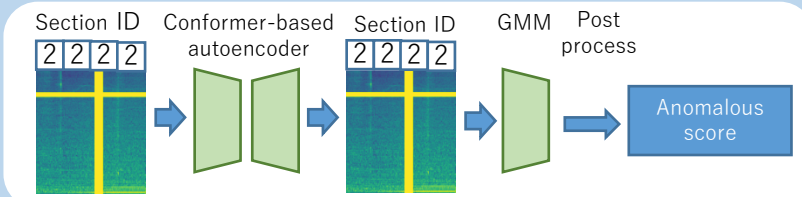
- Ensembled, conformer-based autoencoder (AE)
- Ensembled, binary classifier with metric learning (BC)
- Address the domain shift problem in each method

2. Evaluate with DCASE 2021 Task2

- Our system won in **4th place** in DCASE 2021 Task2

RROPOSED METHOD

✓ **Ensembled, conformer-based autoencoder (AE)**



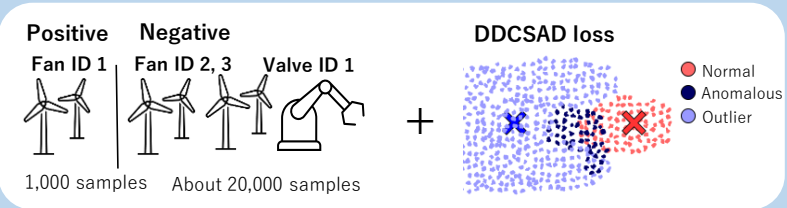
✓ **To boost the autoencoder's performance**

1. Apply SpecAugment and dropout for the input feature sequence
2. Ensemble the model by selecting the N-best models

✓ **To address the domain shift problem**

1. Build separate reconstruction error scoring modules (GMM) for each section and each domain

✓ **Ensembled, binary classifier with metric learning (BC)**



✓ **To boost the autoencoder's performance**

1. Ensemble various models
   - Apply data augmentation, Mixup, ArcFace
   - Select different pseudo-negative sample
   - Use various models (ResNet34, ResNeXt50, EffcientNet-b3)

✓ **To address the domain shift problem**

1. Fine-tuning the model for the target domain
2. Create pseudo-target domain data using Mixup

EXPERIMENTAL EVALUATION

Table 1: *Evaluation results. Values represent the harmonic mean of AUC [%] and pAUC ($p = 0.1$) [%] for each section of each domain. "All/har-mean" column values represent the harmonic mean of AUC and pAUC over all machines, sections and domains.*

		ToyCar		ToyTrain		fan		gearbox		pump		silder		valve		All
eval	Baseline (AE)	61.33	55.63	61.86	**63.26**	57.99	52.54	61.17	**60.58**	56.38	52.55	56.76	51.78	51.22	50.69	56.38
	Baseline (MNV2)	41.81	**57.59**	49.76	43.50	63.65	59.24	53.31	49.55	63.79	64.00	66.17	66.34	53.86	50.86	54.77
	AE ens	54.94	54.95	**65.84**	54.82	62.84	59.00	**66.18**	60.31	58.13	62.14	71.45	62.49	63.30	49.52	59.92
	BC ens	**64.39**	55.07	54.86	52.90	65.97	**63.70**	55.91	50.44	**81.40**	**79.86**	**84.22**	75.69	66.23	**58.49**	63.21
	AE+BC ens (mix)	60.83	56.30	64.64	54.84	**70.01**	63.12	60.93	56.21	70.46	64.84	82.26	**77.36**	**68.78**	55.10	**63.75**
	AE+BC ens (max)	54.94	54.95	**65.84**	54.82	65.97	59.00	55.91	60.31	**81.40**	62.14	**84.22**	62.49	66.23	**58.49**	62.26

- **BC ens** outperformed **AE ens** for machine types pump, slider and valve, regardless of the domain.
  - ✓ The sound is non-stationary (i.e., it includes a variety of intermittent sounds, such as clicks).
- **AE+BC ens (mix)** achieved the best ASD performance.
  - ✓ Using an ensemble of the results from different ASD models focusing on different features contributes to score improvement, since the outputs of the models complement each other.