Conformer-based autoencoder and binary classifier incorporated with metric learning



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 \checkmark

- 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)

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

Evaluate with DCASE 2021 Task2 2.

■ Our system won in 4th place in DCASE 2021 Task2

EXPERIMENTAL EVALUATION

RROPOSED METHOD

Ensembled, conformer-based autoencoder (AE)

Ibuki Kuroyanagi^{1,2}, Tomoki Hayashi^{1,2}, Yusuke Adachi^{1,2}, Takenori Yoshimura¹, Kazuya Takeda², Tomoki Toda²

¹Human Dataware Lab. Co., Ltd. / ²Nagoya University



To boost the autoencoder's performance \checkmark

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

✓ To address the domain shift problem

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

Ensembled, binary classifier with metric learning (BC)



 \checkmark



1.000 samples About 20.000 samples

To boost the autoencoder's performance \checkmark

- 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

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

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
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
	Baseline (AE) Baseline (MNV2) AE ens BC ens AE+BC ens (mix) AE+BC ens (max)	Toy Baseline (AE) 61.33 Baseline (MNV2) 41.81 AE ens 54.94 BC ens 64.39 AE+BC ens (mix) 60.83 AE+BC ens (max) 54.94	ToyCar Baseline (AE) 61.33 55.63 Baseline (MNV2) 41.81 57.59 AE ens 54.94 54.95 BC ens 64.39 55.07 AE+BC ens (mix) 60.83 56.30 AE+BC ens (max) 54.94 54.95	ToyCar ToyT Baseline (AE) 61.33 55.63 61.86 Baseline (MNV2) 41.81 57.59 49.76 AE ens 54.94 54.95 65.84 BC ens 64.39 55.07 54.86 AE+BC ens (mix) 60.83 56.30 64.64 AE+BC ens (max) 54.94 54.95 65.84	ToyCar ToyTrain Baseline (AE) 61.33 55.63 61.86 63.26 Baseline (MNV2) 41.81 57.59 49.76 43.50 AE ens 54.94 54.95 65.84 54.82 BC ens 64.39 55.07 54.86 52.90 AE+BC ens (mix) 60.83 56.30 64.64 54.84 AE+BC ens (max) 54.94 54.95 65.84 54.82	ToyCar ToyTrain fa Baseline (AE) 61.33 55.63 61.86 63.26 57.99 Baseline (MNV2) 41.81 57.59 49.76 43.50 63.65 AE ens 54.94 54.95 65.84 54.82 62.84 BC ens 64.39 55.07 54.86 52.90 65.97 AE+BC ens (mix) 60.83 56.30 64.64 54.84 70.01 AE+BC ens (max) 54.94 54.95 65.84 54.82 65.97	ToyCarToyTrainfanBaseline (AE)61.3355.6361.86 63.26 57.9952.54Baseline (MNV2)41.81 57.59 49.7643.5063.6559.24AE ens54.9454.95 65.84 54.8262.8459.00BC ens 64.39 55.0754.8652.9065.97 63.70 AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.12AE+BC ens (max)54.9454.95 65.84 54.8265.9759.00	ToyCarToyTrainfangearBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.31AE ens54.9454.95 65.84 54.8262.8459.00 66.18 BC ens 64.39 55.0754.8652.9065.97 63.70 55.91AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.93AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.91	ToyCarToyTrainfangearboxBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.55AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.31BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.21AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31	ToyCarToyTrainfangearboxpurBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 56.38Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.5563.79AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.3158.13BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44 81.40 AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.2170.46AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31 81.40	ToyCarToyTrainfangearboxpumpBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 56.3852.55Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.5563.7964.00AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.3158.1362.14BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44 81.4079.86 AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.2170.4664.84AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31 81.40 62.14	ToyCarToyTrainfangearboxpumpsileBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 56.3852.5556.76Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.5563.7964.0066.17AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.3158.1362.1471.45BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44 81.4079.8684.22 AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.2170.4664.8482.26AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31 81.40 62.14 84.22	ToyCarToyTrainfangearboxpumpsilderBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 56.3852.5556.7651.78Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.5563.7964.0066.1766.34AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.3158.1362.1471.4562.49BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44 81.4079.8684.22 75.69AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.2170.4664.8482.26 77.36 AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31 81.40 62.14 84.22 62.49	ToyCarToyTrainfangearboxpumpsildervalBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 56.3852.5556.7651.7851.22Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.5563.7964.0066.1766.3453.86AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.3158.1362.1471.4562.4963.30BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44 81.4079.8684.22 75.6966.23AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.2170.4664.8482.26 77.3668.78 AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31 81.40 62.14 84.22 62.4966.23	ToyCarToyTrainfangearboxpumpsildervalveBaseline (AE)61.3355.6361.86 63.26 57.9952.5461.17 60.58 56.3852.5556.7651.7851.2250.69Baseline (MNV2)41.81 57.59 49.7643.5063.6559.2453.3149.5563.7964.0066.1766.3453.8650.86AE ens54.9454.95 65.84 54.8262.8459.00 66.18 60.3158.1362.1471.4562.4963.3049.52BC ens 64.39 55.0754.8652.9065.97 63.70 55.9150.44 81.4079.8684.22 75.6966.23 58.49 AE+BC ens (mix)60.8356.3064.6454.84 70.01 63.1260.9356.2170.4664.8482.26 77.3668.78 55.10AE+BC ens (max)54.9454.95 65.84 54.8265.9759.0055.9160.31 81.40 62.14 84.22 62.4966.23 58.49

- 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 \checkmark focusing on different features contributes to score improvement, since the outputs of the models complement each other.

40