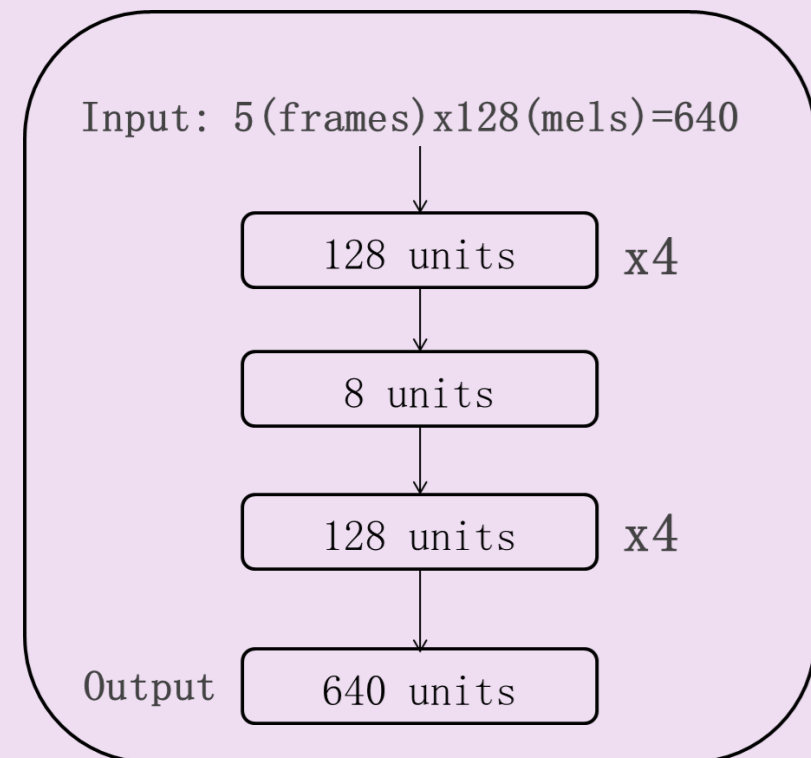


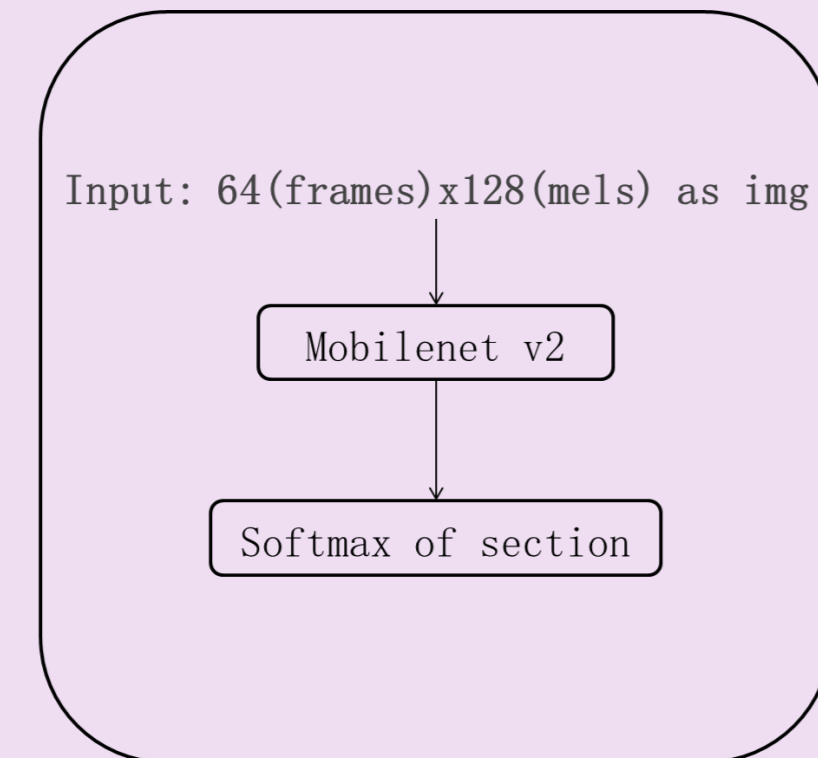
Introduction

Two mainstream method

unsupervised AE-based model



supervised CNN-based model



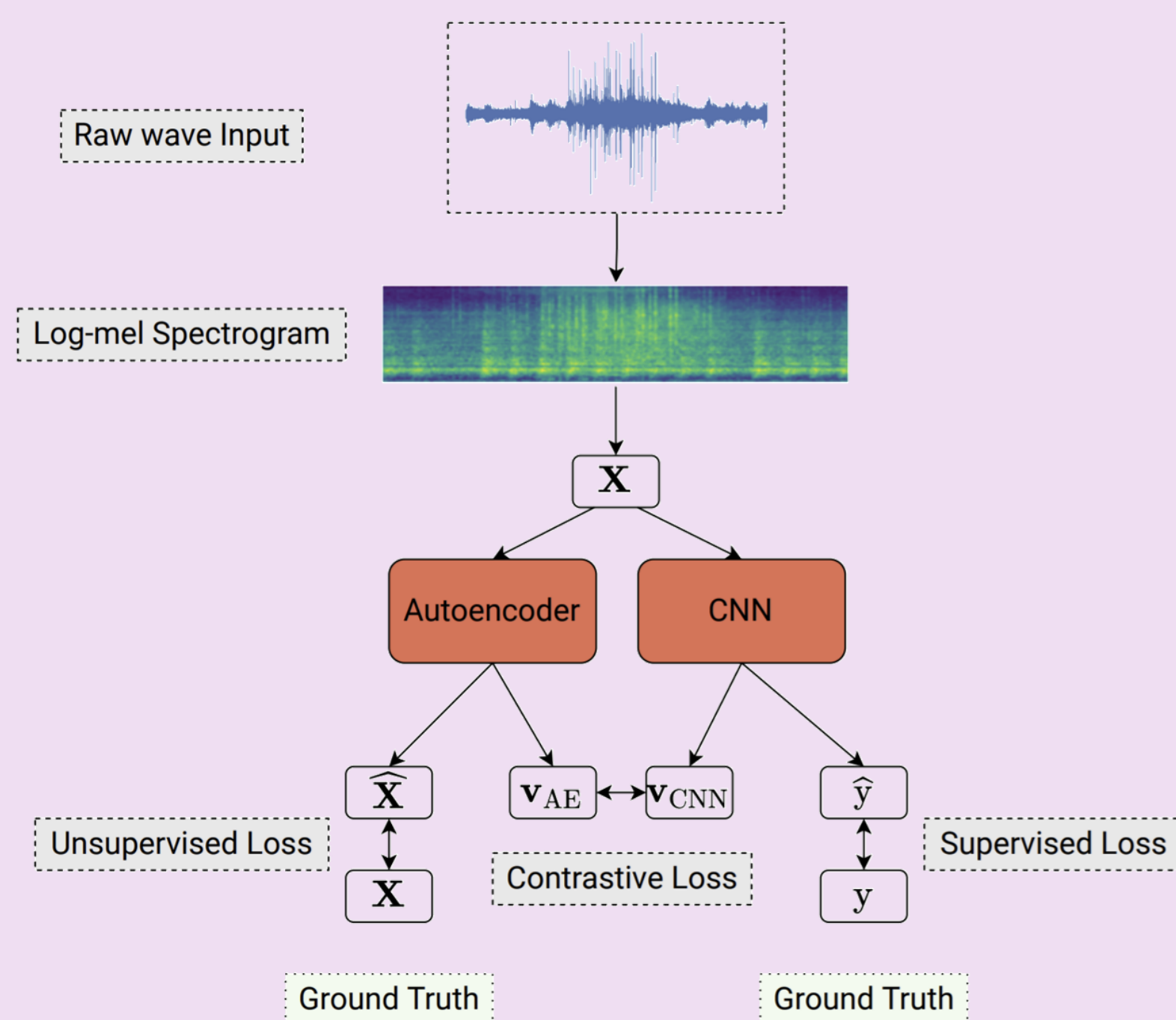
Contrastive Learning

Obtain two different representations of a single sample:

$$\mathbf{p} = \mathbf{v}_{AE} \quad \mathbf{u} = \mathbf{v}_{CNN}$$

Calculate the contrastive loss

$$\mathcal{L}_{contrastive}(\cdot) = - \sum_i \log \frac{\exp(\langle \mathbf{u}_i, \mathbf{p}_i \rangle / \rho)}{\sum_{j \neq i} \exp(\langle \mathbf{u}_i, \mathbf{p}_j \rangle / \rho)}$$



Data Augmentation

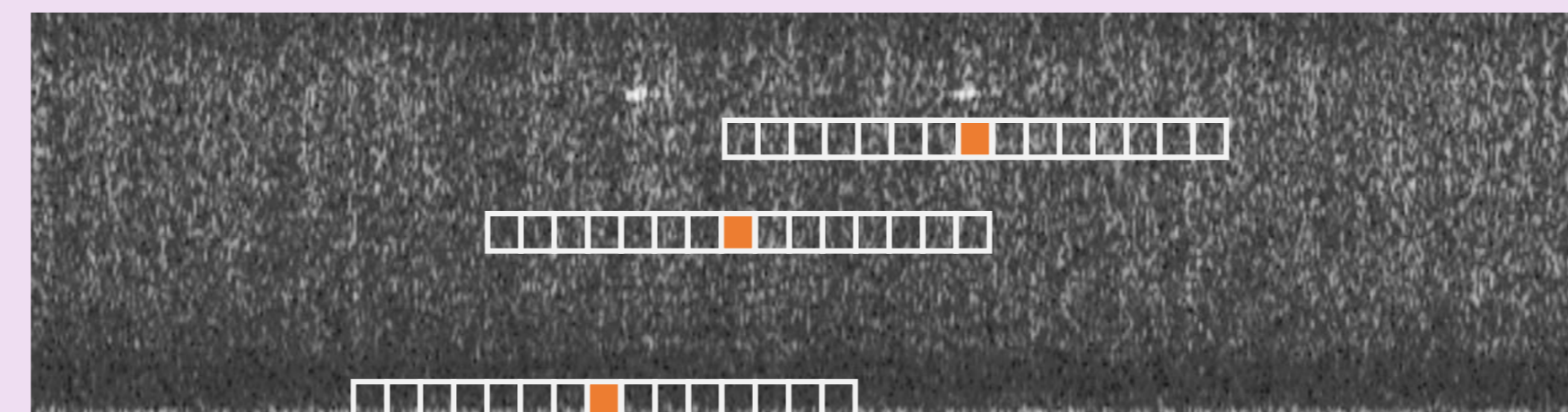
Conventional techniques:

Mixup, Time masking, Frame-shifting

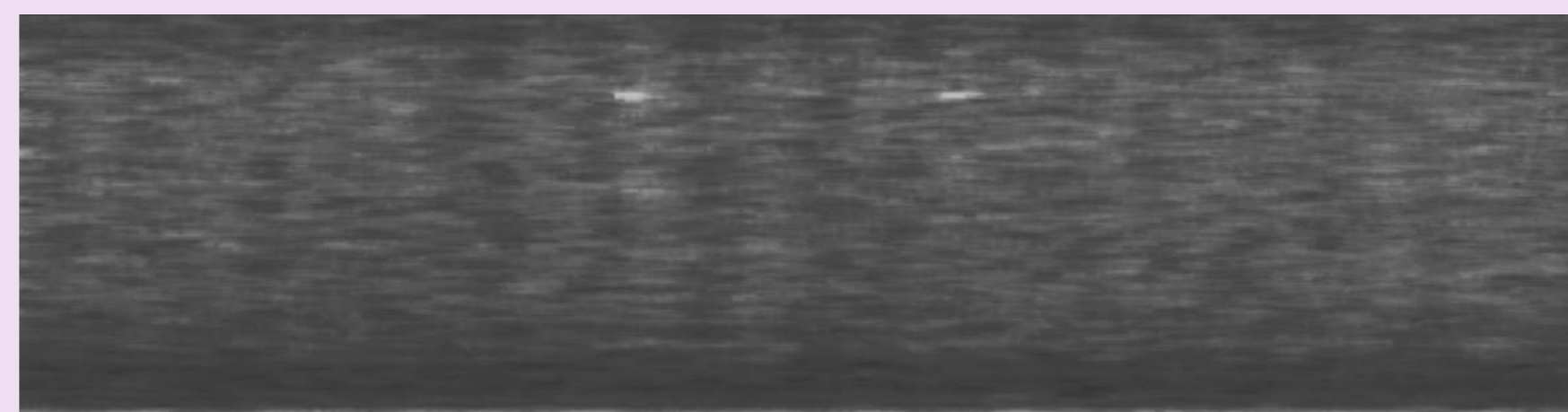
Explore other techniques:

We use a **median filter** to suppress short-time noise

Original



Processed



Challenge Results

For the challenge, our method ranked 9th out of 27 participated methods

Model	Official Score	Fan	Gearbox	Slider	Toy Train	Toy Car	Pump	Valve
AE Baseline	56.375	60.68	65.49	57.22	68.51	65.93	58.30	51.87
MBv2 Baseline	54.770	64.96	51.14	72.92	42.91	42.73	67.97	53.13
1st	66.798	61.01	63.07	83.18	69.15	75.27	86.76	65.36
2nd	64.956	86.48	67.45	83.05	45.60	60.88	85.04	71.49
3rd	64.201	88.98	57.75	86.84	57.50	69.83	74.82	62.74
4th	63.745	66.60	62.53	86.27	61.79	61.70	74.60	62.36
5th	62.593	68.98	67.74	79.88	61.71	73.32	71.87	63.73
6th	62.239	82.65	57.20	83.76	53.43	58.67	85.54	60.54
7th	61.480	87.68	56.56	76.66	48.24	70.60	72.54	60.70
8th	61.186	73.17	64.70	69.89	51.71	68.23	78.65	53.93
Ours best	60.966	90.68	58.00	77.34	47.49	53.81	77.82	53.53

Our method performed best on the Fan dataset, especially from the perspective of **pAUC** metric

Model	Fan (AUC)	Fan (pAUC)
AE Baseline	60.68	50.50
MBv2 Baseline	64.96	58.14
2nd	90.22	71.19
3rd	88.98	70.20
4th	88.09	70.84
Ours	90.68	79.99

Development Dataset Results

Dcase 2021 Task 2 Dataset

“+CL” represents adding contrastive learning

“+MF” represents the addition of median filtering

Model	Fan	Gearbox	Slider	Toy Train	Toy Car	Pump	Valve	Score
MBv2	60.30	57.43	59.43	51.10	53.60	56.17	55.19	56.01
+ CL	60.61	58.87	60.70	50.92	52.51	56.90	54.38	56.18
+ MF	64.08	65.38	59.83	49.69	55.38	59.50	53.74	57.75
+ CL, MF	64.45	67.16	58.66	51.89	56.15	57.27	53.46	57.99

Conclusion

- Our proposed training framework exceeds the baseline model for some machine types, while no additional parameters are introduced during inference.
- Appropriate data augmentation technology can greatly improve the performance of anomaly sound detection system, and further investigation is needed.