NOISE-AWARE FAR-CONDITIONED MASKED SPECTROGRAM MODELING WITH DSP FEATURE FUSION

Technical Report

Iljoo Jeong[*]

Korea Electronics Technology Institute (KETI)
Seongnam, South Korea
iljjeong@keti.re.kr

ABSTRACT

We address first-shot, noise-aware anomalous sound detection under source and target domain shift, using the synchronized Near and Far (close and distant) stereo recordings, and we train the entire system from scratch without any pretrained models or external data. Our design is motivated by an invariance trap that we observe in this setting: representations learned from scratch with discriminative or denoising objectives, such as machine-ID classification, pseudo-cluster metric learning, and far-to-near denoising, collapse fault sensitivity and reach near-chance detection, because one-class training provides no gradient toward fault-relevant directions. To avoid this, we combine two complementary branches. First, we preserve fault-relevant time-frequency structure with windowed masked spectrogram modeling (MSM), namely a convolutional encoder-decoder that is trained per machine to reconstruct 60%-masked 2.5 s log-mel patches from a bottleneck embedding; the embedding itself, rather than its reconstruction residual, is used for scoring. Second, we inject fault-axis sensitivity that learning alone does not discover, using non-neural DSP descriptors computed directly from each Near and Far window, i.e., a far-context ridge-residual descriptor and a temporal modulation descriptor. From normal Near and Far pairs we estimate a Far-correlated shared nuisance subspace, and we form a Far-conditioned MSM feature that attenuates this shared operational and environmental variation while preserving Near-specific variation. The three blocks are block-balanced, concatenated, and scored by a single domain-conditional kNN density estimator with target-domain shrinkage, and the window scores are aggregated by a top-10% mean into one anomaly score per file. A single scalar that controls the Far-conditioning is selected on the development machines and fixed for evaluation, so that no per-machine tuning is applied and no anomaly labels are used during training or model selection.

1. REFERENCES

[1] T. Nishida, N. Harada, D. Takeuchi, D. Niizumi, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and Discussion on DCASE 2026 Challenge Task 2: Noise-aware Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," arXiv e-prints: 2606.01578, 2026.

[2] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: Another Dataset of Miniature-Machine Operating Sounds for Anomalous Sound Detection under Domain Shift Conditions," in *Proc. DCASE Workshop*, 2021, pp. 1–5.

[3] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "MIMII DG: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task," in *Proc. 7th DCASE Workshop (DCASE2022)*, 2022.

[4] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, "First-Shot Anomaly Detection for Machine Condition Monitoring: A Domain Generalization Baseline," in *Proc. 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 191–195.

[*]External resources: none (trained from scratch, no pretrained models, no external data).