FIRSTSHOT UNSUPERVISED ANOMALOUS SOUND DETECTION USING AUTOENCODERS AND GAMMATHRESHOLDING

Technical Report

*Meghan Kret**

The Cooper Union
Albert Nerkin School of Engineering
New York, NY, USA
meghan.kret@cooper.edu

ABSTRACT

We introduce a two-stage autoencoder baseline for DCASE 2025 Task 2. Stage 1 pre-trains an autoencoder exclusively on the seven development machines to learn domain-general normal-sound features. Stage 2 fine-tunes that model on each evaluation machines 990 source and 10 target normal clips, then infers on its 200 unlabeled test clips. Anomaly scores are mean-squared reconstruction errors; a Gamma distribution fit to the fine-tune training errors supplies a fixed 90th-percentile threshold. The pipeline satisfies the first-shot requirement (no per-machine hyper-parameter tuning) while leveraging dev data for improved generalization.

***Index Terms*—** anomalous sound detection, first-shot, autoencoder, pre-training, fine-tuning

1. INTRODUCTION

Unsupervised anomalous sound detection (ASD) aims to detect unseen machine failures using only normal-sound training data. DCASE 2025 Task 2 further imposes (i) domain-shift robustness and (ii) first-shot generalization to new machine types [1]. We extend the official autoencoder (AE) baseline of [2] with a development-set *pre-training* stage followed by per-machine *fine-tuning*, enabling knowledge transfer without violating first-shot constraints.

2. METHOD

2.1. Feature Extraction

Audio is resampled to 16kHz. We compute 128-band log-Mel spectrograms using 64ms frames and 50 % hop, giving $T \approx 309$ frames for a 10s clip. Five consecutive frames ($P = 5$) are concatenated into one 640-dimensional vector.

2.2. Model

A fully-connected AE is used:

$$640 \to 128 \to 64 \to 8 \text{ (bottleneck)} \to 64 \to 128 \to 640,$$

with BatchNorm + ReLU after each hidden layer. Training loss is mean-squared error (MSE).

*Student at The Cooper Union, Albert Nerkin School of Engineering.

2.3. Stage 1 Pre-training on Development Data

Normal windows from all seven development machines (`dev_data/raw/*/train`) are pooled. The AE is trained for 20 epochs (batch 64, Adam 0.001). Although no machine-specific thresholds are retained from this phase, the encoder acquires domain-agnostic representations useful for subsequent fine-tuning.

2.4. Stage 2 Fine-tuning and Inference

For each of the eight evaluation machines:

1. **Fine-tune.** The pretrained weights initialize the AE, which is then updated for 20 epochs on that machines 1 000 normal clips (990 source + 10 target).
2. **Threshold.** Reconstruction errors on the fine-tune training windows are modeled by a Gamma distribution $\mathrm{Gamma}(a, \mathrm{scale}; \mathrm{loc} = 0)$. The decision threshold is the 90th percentile:

$$\tau = F^{-1}_{\mathrm{Gamma}}\big(0.90; a, 0, \mathrm{scale}\big).$$

3. **Inference.** For each 10s test clip, the anomaly score $A_\theta(X) = \frac{1}{DT}\sum_t \big\|\xi_t - r_\theta(\xi_t)\big\|_2^2$ is compared with τ to yield a binary decision. Scores and decisions are written to the required CSV files.

3. EQUATIONS

Using log-Mel frames $X_t \in R^F$ and context length P:

$$\xi_t = [X_t^\top \, X_{t+1}^\top \dots X_{t+P-1}^\top]^\top \in R^D, \quad D = P\,F. \tag{1}$$

$$A_\theta(X) = \frac{1}{DT}\sum_{t=1}^{T} \|\xi_t - r_\theta(\xi_t)\|_2^2. \tag{2}$$

Threshold: $\tau = F^{-1}_{\mathrm{Gamma}}(0.90; a, 0, \mathrm{scale})$.

4. DISCUSSION

Pre-training supplies a generic embedding of normal machine sounds; fine-tuning adapts this embedding with only 1 000 clips per new machine, preserving first-shot validity. Compared with training each machine from scratch, we observe faster convergence and more stable detection under target-domain noise (numerical results omitted for brevity).

5. CONCLUSION

We presented a simple pre-train+fine-tune strategy for first-shot ASD. The method adheres strictly to Task 2 rules, requires no anomaly examples, and leverages development data without per-machine hyper-parameter tuning.

6. ACKNOWLEDGMENT

The author thanks Professor Sam Keene for guidance.

7. REFERENCES

[1] DCASE 2025 Task 2, `https://dcase.community/challenge2025/`, 2025.

[2] N. Harada *et al.*, First-shot anomaly detection for machine condition monitoring, *Proc. EUSIPCO*, 2023.

[3] N. Harada *et al.*, ToyADMOS2, *Proc. DCASE Workshop*, 2021.

[4] K. Dohi *et al.*, MIMII-DG dataset, *Proc. DCASE Workshop*, 2022.