PA-LORA: PROTOTYPE ANCHORED LOW-RANK ADAPTATION FOR DOMAIN-INCREMENTAL AUDIO CLASSIFICATION

Technical Report

Bohan Hu, Yiqiang Cai, Shengchen Li

Xi'an Jiaotong-Liverpool University
School of Advanced Technology, Suzhou, China
{bohan.hu24, yiqiang.cai21}@student.xjtlu.edu.cn, shengchen.li@xjtlu.edu.cn

ABSTRACT

The Task 7 of DCASE Challenge 2026 focuses on developing a universal domain-incremental learning system that learns to classify audio from different domains sequentially over time without significantly forgetting the knowledge of any of the previously learned domains. This technical report details the systems we submitted. Rather than rebuilding decision boundaries for each new domain, we propose Prototype Anchored Low-Rank Adaptation (PA-LoRA), which anchors the basic classifier as a frozen prototype space and learns only lightweight, low-rank boundary deformations for new domains, with a learnable gate that autonomously calibrates adaptation. Continual Normalization combines group and batch normalization to prevent feature statistics drift. Furthermore, we introduce negative sampling-based domain routing regularization to reduce misrouting errors during incremental training. On the DIL-DCASE26 development dataset, our best system LoRA-NS achieves an average domain-agnostic accuracy of 61.51%.

***Index Terms*—** domain-incremental learning, continual learning, audio classification, catastrophic forgetting, batch normalization, low-rank adaptation

1. INTRODUCTION

Deep learning models for audio classification are conventionally trained on static datasets covering all expected acoustic conditions. In practice, however, models encounter new recording environments, devices, or sound sources (collectively, *domains*) over their operational lifetime. Continually retraining from scratch is computationally prohibitive and often infeasible due to storage or privacy constraints. This motivates domain-incremental learning (DIL): a model must sequentially learn to classify sounds from new domains without revisiting data from earlier ones.

The DCASE 2026 Challenge Task 7 formalizes this problem [1]. As shown in Figure 1, a single classifier must learn 10 sound classes from three domains (D1, D2, D3) revealed sequentially, with no access to previous domains' data during incremental training.

The central obstacle is catastrophic forgetting [3, 4]: fine-tuning exclusively on new domain data steers parameters toward a new optimum, overwriting previously learned knowledge. The official baseline [2] partially mitigates this by freezing convolutional layers after D1 and using domain-specific batch normalization (BN) layers. However, it remains vulnerable to BN statistics drift and offers no mechanism to adapt the classifier without interfering with past knowledge.

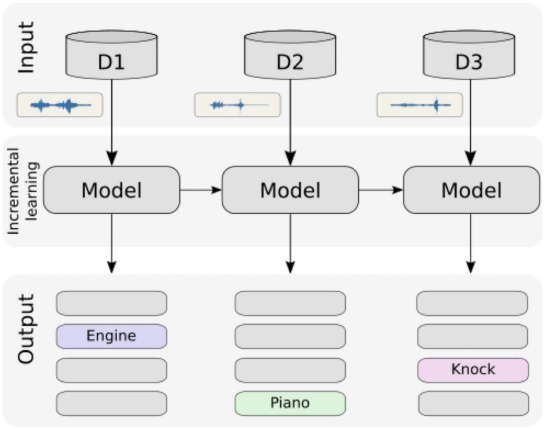

Figure 1: *Overview of the Domain-Agnostic Incremental Learning for Audio Classification task.*

We propose a system built on a geometric insight: adapting to a new domain requires *bending*, not *rebuilding*, the decision boundaries learned from previous domains. We freeze the D1 classifier as an anchored prototype space and introduce low-rank additive updates that apply controlled, low-dimensional deformations. This strategy preserves the global class structure while enabling domain-specific adjustments.

2. DATA PREPROCESSING AND AUGMENTATION

2.1. Dataset

We use the DIL-DCASE26 dataset containing 10 sound classes from three domains [1]. The development set provides labeled audio for D2 and D3. D1 knowledge is embedded in the supplied baseline checkpoint. Evaluation uses the development test split covering all three domains.

2.2. Feature Extraction

All audio is resampled to 32 kHz and segmented or padded to 4 seconds. We extract 64-band log-mel spectrograms using a Hamming window of 1024 samples, a hop length of 320 samples, and frequency limits of $f_{\min} = 50$ Hz and $f_{\max} = 14$ kHz.

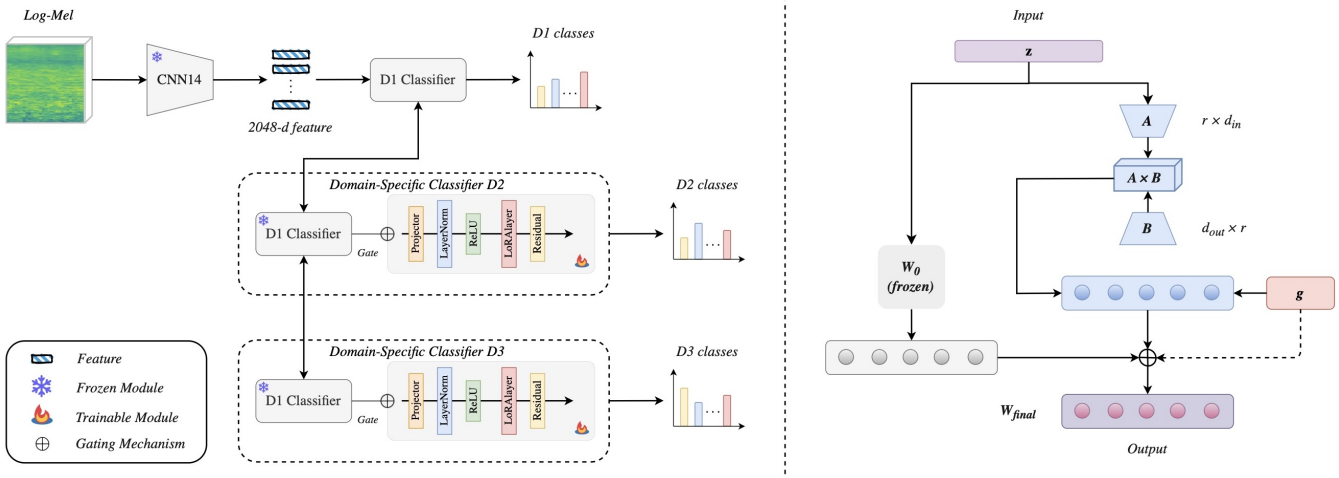

Figure 2: *Overview of the proposed system. Left: Complete architecture with domain-specific PA-LoRA adapters (D2/D3) anchored to the frozen D1 classifier. Right: Detailed view of the PA-LoRA computation, where the frozen prototype weight matrix* $\mathbf{W}_0$ *is augmented by a low-rank deformation field* $\mathbf{B}, \mathbf{A}$ *scaled by a learnable gate* g.

2.3. Data Augmentation

Data augmentation is critical in domain-incremental learning, where each domain is encountered only once and the model cannot revisit past data. We employ three plug-and-play augmentation strategies to improve robustness and prevent overfitting to domain-specific patterns.

- **SpecAugment** [7] applies random frequency and time masking to the input mel-spectrograms. Contiguous bands of frequency bins and time frames are masked (10% mask size, $p = 0.8$ for each mask type), forcing the model to rely on partial and diverse acoustic cues rather than memorizing specific spectro-temporal patterns.
- **Freq-MixStyle** [8] mixes frequency-wise statistics between two samples. Given frequency-wise means μ_f, μ'_f and variances σ_f, σ'_f, a new sample is generated via:

$$\mu_{\text{mix}} = \lambda\mu_f + (1-\lambda)\mu'_f, \quad \sigma_{\text{mix}} = \lambda\sigma_f + (1-\lambda)\sigma'_f, \quad (1)$$

  with $\lambda \sim \text{Beta}(\alpha, \alpha)$. This perturbation in the spectral statistics domain improves generalization across different acoustic conditions.
- **Soft Mixup** [9] linearly interpolates pairs of training samples:

$$\mathbf{x}_{\text{new}} = \lambda\mathbf{x}_i + (1-\lambda)\mathbf{x}_j, \quad (2)$$

  with labels mixed correspondingly. We use $\alpha = 0.3$, producing conservative mixing that preserves recognizable class characteristics while smoothing decision boundaries. The loss is computed as:

$$\mathcal{L} = \lambda\mathcal{L}_{\text{CE}}(\hat{\mathbf{y}}, \mathbf{y}_i) + (1-\lambda)\mathcal{L}_{\text{CE}}(\hat{\mathbf{y}}, \mathbf{y}_j), \quad (3)$$

  applied with probability $p = 0.5$.

All augmentations are applied sequentially during training. SpecAugment and Freq-MixStyle operate on the spectrogram representations, while Soft Mixup operates at the batch level.

3. PROPOSED METHOD

Our system is built on the insight that the D1 classifier learns a universal class-prototype space capturing fundamental acoustic relationships among the 10 categories. Rather than rebuilding this space for each new domain—which would overwrite learned boundaries—we anchor the D1 classifier and learn only low-rank residual offsets. The complete architecture is shown in Figure 2.

3.1. Continual Normalization

While classification boundaries are anchored to the D1 prototype space, the features feeding into them must remain stable across domains. Standard BN is vulnerable: training on a new domain t shifts running statistics toward the new distribution via:

$$\mu_{\text{running}} \leftarrow \alpha\mu_{\text{batch},t} + (1-\alpha)\mu_{\text{running}}, \quad (4)$$

causing features from earlier domains to be mis-normalized. The prototype boundaries remain intact, but features drift away from them.

We adopt Continual Normalization (CN) [5, 10], which interleaves Group Normalization (GN) and BN:

$$\text{CN}(\mathbf{x}) = \text{BN}(\text{GN}(\mathbf{x})). \quad (5)$$

GN removes domain-specific style variations via per-group normalization; subsequent BN operates on more domain-invariant representations, yielding stable running statistics.

Every BN layer in the CNN14 backbone is replaced with CN. Each domain t retains its own CN parameters (γ_t, β_t and running statistics), while shared convolutional kernels remain frozen after D1. Only the t-th CN parameters are updated when learning domain t, structurally isolating domain-specific normalization from the shared feature extractor.

3.2. Prototype Anchored Low-Rank Adaptation

With features stabilized by CN, we address the classifier. Standard fine-tuning updates:

$$\mathbf{W}_{\text{final}} = \mathbf{W}_0 - \eta \nabla \mathcal{L}_t, \tag{6}$$

deforming decision boundaries globally and erasing earlier knowledge. Our Prototype Anchored Low-Rank Adaptation (PA-LoRA) is grounded in the geometric insight that domain shift requires *deforming*, not *rebuilding*, boundaries.

Given the frozen D1 weight matrix $\mathbf{W}_0$, we learn a low-rank deformation field:

$$\mathbf{W}_{\text{final}} = \mathbf{W}_0 + g \cdot \frac{\alpha}{r} \cdot \mathbf{BA}, \tag{7}$$

where $\mathbf{A} \in R^{r \times d_{\text{in}}}$, $\mathbf{B} \in R^{d_{\text{out}} \times r}$, $r \ll \min(d_{\text{in}}, d_{\text{out}})$, and $g \in [0, 1]$ is a learnable scalar gate. The rank $r = 4$ configuration reduces trainable parameters by over 99%, yet suffices to capture the necessary boundary corrections—confirming that domain shift operates in a low-dimensional subspace.

The gate g autonomously learns the optimal degree of deformation: when $g \to 0$, the model relies on the frozen prototype boundaries; when $g \to 1$, the full deformation is applied. Gradients push g upward when the LoRA update reduces loss, and downward when it causes interference.

For each new domain $t \in \{2, 3\}$, we create an independent adapter with a projector (2048→256), the Gated LoRA layer, and a residual connection. Only the current domain's adapter is trained; all others remain frozen, ensuring D3 training never alters D2-specific corrections.

3.3. Domain Routing with Negative Sampling

At test time, the model must route each sample to the appropriate domain-specific pathway without domain labels. We adopt entropy-based selection from the baseline [2]: the pathway with lowest predictive entropy is chosen. However, a critical failure mode arises: the D3 classifier may learn to output confident predictions on D2 samples sharing similar acoustic patterns, causing misrouting and degrading D2 accuracy even though the D2 classifier remains unchanged.

To address this, we introduce negative sampling during D3 training. A small replay buffer of D2 samples (2,000 examples, populated during D2 training) is maintained. During each D3 batch, we sample D2 examples and penalize overconfident predictions via a uniform-entropy loss:

$$\mathcal{L}_{\text{uniform}} = \left(\log K - H(\mathbf{p}_{\text{D3}}(\mathbf{x}_{\text{D2}}))\right)^2, \tag{8}$$

where $K = 10$ and $H(\cdot)$ is the predictive entropy. The total D3 loss becomes:

$$\mathcal{L}_{\text{D3}} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{uniform}}. \tag{9}$$

We evaluate two variants: **Uniform NS** penalizes all D2 samples equally; **Adaptive NS** applies the penalty only when D3 entropy falls below a threshold τ, preventing over-regularization of the D3 classifier on already well-handled samples. Critically, the buffer requires no D2 labels—only the domain origin—maintaining compliance with the challenge's data access constraints.

Model	D2*	D2	D3	Avg.
Baseline	58.60	59.00	46.10	52.50
LoRA-G	68.86	66.35	54.47	60.41
LoRA-ANS	70.89	66.20	**55.90**	60.64
LoRA-NS	**71.05**	**69.17**	53.85	**61.51**

Table 1: *Domain-agnostic evaluation accuracy (%) on the development test set after incremental training. D2* results are calculated after D2 training. D2: accuracy on D2 after completing D3 training. Avg.: unweighted average of D2 and D3 domain-wise accuracies.*

4. EXPERIMENTAL SETUP

We follow the official three-stage incremental protocol [2]. Official D1 pretrained weights are loaded into our CN backbone, with BN parameters mapped to the BN component of corresponding CN layers and the D1 classifier loaded directly. All parameters remain frozen after Stage 1. In Stage 2 (learn D2), we unfreeze only D2-indexed CN layers and the D2 PA-LoRA adapter, training for 150 epochs using Adam with initial learning rate 5×10^{-4} and cosine annealing scheduler that decays to 1% of the initial value. The batch size is set to 16. In Stage 3 (learn D3), all D2-specific parameters are frozen, and D3 CN layers and PA-LoRA adapter are unfrozen and trained identically, except the initial learning rate is increased to 1×10^{-3}. For data augmentation, α of Soft Mixup is set to 0.3 with probability $p = 0.5$, and α and p of Freq-MixStyle are set to 0.3 and 0.7, respectively. SpecAugment uses a mask size of 10% with probability $p = 0.8$ for each masking type. For negative sampling, System 2 (LoRA-NS) uses uniform penalty weight $\lambda = 0.5$ with 30% negative ratio, while System 3 (LoRA-ANS) uses adaptive penalty weight $\lambda = 0.04$ with threshold $\tau = 0.8$ and 20% negative ratio. Evaluation follows the official metric: per-class accuracy is averaged within each domain, and the three domain-wise averages are averaged to produce the overall score, ensuring equal domain weight regardless of sample count.

5. SUBMISSION

We submitted a total of three systems, all sharing the same CNN14 backbone with Continual Normalization and PA-LoRA adapters. The systems differ only in the domain routing regularization strategy applied during D3 training, as summarized in Table 1.

System 1 (LoRA-G) serves as our base PA-LoRA system without any routing regularization, relying solely on entropy-based domain selection inherited from the baseline. **System 2 (LoRA-ANS)** employs adaptive negative sampling, where the penalty is applied only when D3 entropy falls below τ, preserving D3 discriminative capacity while still correcting routing errors. **System 3 (LoRA-NS)** extends this with uniform negative sampling, maintaining a replay buffer of 2,000 D2 samples and applying a uniform-entropy penalty to all D2 examples to suppress overconfident predictions on out-of-domain data. Among the three submissions, System 3 achieves the best average accuracy and is recommended as the primary system for scenarios prioritizing balanced domain performance, while System 2 is preferable when maximizing accuracy on the most recently learned domain is the primary objective.

6. CONCLUSIONS

We presented a domain-incremental audio classification system integrating Continual Normalization, Prototype-Anchored Low-Rank Adaptation (PA-LoRA), and negative sampling-based routing regularization. CN prevents BN statistics drift by interleaving group and batch normalization. PA-LoRA anchors the D1 classifier as a prototype space and learns only low-rank boundary deformations via a gated mechanism. Negative sampling corrects domain routing errors during incremental training. On the DIL-DCASE26 dataset, our best system LoRA-NS achieves 61.51% average D2-D3 accuracy (vs. 52.50% baseline), with D2 retention reaching 69.17% and D3 accuracy 53.85%. Our approach adds negligible parameters ($<$1% of total per domain) and maintains the baseline's inference cost. Future directions include dynamic adapter allocation for open-ended domain sequences and learned domain detection to replace entropy-based routing.

7. REFERENCES

[1] R. Casciotti, M. Mulimani, M. Harju, J. R. Jensen, and A. Mesaros, "Domain-agnostic incremental learning for sound classification: A DCASE 2026 challenge task," 2026, *arXiv:2606.02173*.

[2] M. Mulimani and A. Mesaros, "Domain-incremental learning for audio classification," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[3] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 128–135, 1999.

[4] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," in *Psychology of Learning and Motivation*, vol. 24, Academic Press, 1989, pp. 109–165.

[5] Q. Pham, C. Liu, and S. Hoi, "Continual normalization: Rethinking batch normalization for online continual learning," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.

[6] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *Proc. International Conference on Learning Representations (ICLR)*, 2022.

[7] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A simple data augmentation method for automatic speech recognition," in *Proc. Interspeech*, 2019, pp. 2613–2617.

[8] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Domain generalization with MixStyle," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.

[9] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. International Conference on Learning Representations (ICLR)*, 2018.

[10] M. Mulimani and A. Mesaros, "A closer look at class-incremental learning for multi-label audio classification," *IEEE Trans. Audio, Speech, Language Process.*, vol. 33, pp. 1293–1306, 2025.