ROUTER-BASED PROGRESSIVE KNOWLEDGE REUSE FOR DOMAIN-AGNOSTIC INCREMENTAL AUDIO CLASSIFICATION

Technical Report

HaeChun Chung

Republic of Korea
hc.chung@kt.com

ABSTRACT

Domain-Incremental Learning (DIL) for sound classification requires models to learn new domains without forgetting previously acquired knowledge. In DCASE 2026 Task 7, this challenge is further complicated by the absence of domain identity during inference. To address this problem, we propose a Router-based Progressive Knowledge Reuse framework that introduces domain-specific classifiers and lightweight routing modules for adaptive knowledge reuse across domains. An entropy-threshold strategy is employed during training to regulate the transition between knowledge reuse and knowledge expansion, while inference is performed through entropy-based selection of progressively aggregated domain predictions. Experiments on the DCASE 2026 Task 7 development set show that the proposed method achieves 72.14% accuracy after learning Domain 2 and 60.7% average accuracy after learning Domain 3, outperforming the reproduced baseline by 12.72 and 7.23 percentage points, respectively. The results demonstrate that the proposed approach effectively mitigates forgetting while improving domain-agnostic incremental learning performance.

Index Terms— Domain-Agnostic Incremental Learning, Knowledge Reuse

1. INTRODUCTION

The ability of machine learning systems to continuously acquire knowledge from evolving data streams without forgetting [1] previously learned information is a fundamental challenge in artificial intelligence. In practical sound applications, a deployed model must adapt over time to new acoustic domains—such as sounds recorded in different environments or under varying acoustic conditions—while retaining its classification capability on all previously encountered domains. This problem, known as Domain-Incremental Learning (DIL), is highly critical since data from past domains is typically unavailable due to storage or privacy constraints. The DCASE 2026 Task 7 [2], Domain-Agnostic Incremental Learning for Sound Classification, further extends this challenge into a more rigorous and realistic scenario. In this task, a model must sequentially learn ten target sound classes across three distinct domains without access to prior domain data. Crucially, the system is required to classify sound events during inference without any explicit domain identity (domain-agnostic), where mitigating catastrophic forgetting is essential to prevent severe performance degradation across the entire domain sequence.

Table 1: Accuracy (%) of the baseline and proposed systems across domains D2 and D3 on the development test split. ET denotes the entropy threshold used during router supervision. Smaller ET values enforce stricter confidence requirements for reusing previously learned domain knowledge.

System	After learning D2		After learning D3		
	D2	Avg	D2	D3	Avg
Baseline [2]	59.42	59.42	59.23	47.72	53.47
System1 (No ET)	69.47	69.47	67.34	50.77	59.06
System2 (ET=0.1)	70.53	70.53	66.83	51.16	58.99
System3 (ET=0.05)	71.02	71.02	67.61	**51.53**	59.57
System4 (ET=0.01)	**72.14**	**72.14**	**70.03**	51.37	**60.7**

Prior work on Domain-Incremental Learning (DIL) for sound classification has extensively explored architecture-based methods, which partition model parameters into domain-shared and domain-specific components to mitigate forgetting. The baseline system provided for this challenge adopts this paradigm by integrating domain-specific Batch Normalization (DS-BN) layers within PANNs CNN14 [3] architecture while keeping the remaining network parameters shared across domains. [2, 4, 5] During inference, since explicit domain information is unavailable, the baseline estimates the most appropriate domain by selecting the domain branch that produces the lowest prediction entropy. However, this design presents two limitations. First, although DS-BN layers enable domain-specific feature adaptation, the use of a single shared classifier constrains all domains to utilize the same classification head, potentially limiting the model's ability to capture domain-specific decision patterns. Second, the inference procedure relies exclusively on the prediction from a single selected domain branch. As a result, knowledge learned from multiple domains cannot be jointly exploited during inference, reducing the potential benefits of cross-domain knowledge transfer. Consequently, inaccurate domain selection under ambiguous domain conditions may cause classification errors and degrade overall performance.

To address this limitation, we propose a Router-based Progressive Knowledge Reuse framework for domain-agnostic incremental audio classification. Instead of relying solely on entropy-based domain selection, the proposed method introduces domain-specific classifiers and lightweight routing modules that learn how much knowledge should be reused from previously learned domains. During incremental learning, router networks are trained to identify the minimum knowledge depth required to correctly classify each sample, while classifier outputs from multiple domains are combined through router-guided weighting. Furthermore, an entropy-threshold strategy is introduced during training to control the transition between knowledge reuse and knowledge expansion. Specifically, when predictions derived from previously learned domain classifiers exhibit sufficiently low entropy, the router is encouraged to retain and exploit the existing knowledge. In contrast, when the confidence of previously acquired classifiers is inadequate, the router is trained to incorporate the newly introduced domain-specific classifier, facilitating adaptive collaboration between previously learned and newly acquired knowledge representations. Experimental results on the DCASE 2026 Task 7 development set demonstrate the effectiveness of the proposed approach. After learning Domain 2, the proposed system achieves 72.14% accuracy compared to 59.42% for the baseline. After learning Domain 3, the proposed method attains 70.03% accuracy on Domain 2 and 51.37% accuracy on Domain 3, yielding an average accuracy of 60.7%, which exceeds the baseline average accuracy of 53.47% by 7.23 percentage points. These results indicate that the proposed routing mechanism provides a more reliable utilization of domain knowledge while maintaining strong incremental learning performance.

2. METHOD

2.1. Model Architecture

The proposed system builds upon the PANNs CNN14-based architecture provided in the challenge baseline, preserving all convolutional layers and domain-specific Batch Normalization (DS-BN) parameters to enable direct inheritance of the pretrained D1 checkpoint. Two architectural modifications are introduced.

Domain-Specific Classifiers. While the baseline employs a single shared fully-connected (FC) classification head across all domains, we replace it with a set of domain-specific classifiers $\{C_d\}_{d=1}^{D}$, where each C_d is an independent linear layer mapping the 2048-dimensional shared feature representation $f \in \mathbb{R}^{2048}$ to K target classes. The D1 classifier is initialized from the pretrained checkpoint, while classifiers for subsequent domains are randomly initialized. When incrementally training on domain d, only classifier C_d is set as trainable, while all other classifiers $\{C_{d'}\}_{d' \neq d}$ are frozen.

Router Modules. We introduce a lightweight Router module $\{R_d\}_{d=1}^{D}$ for each domain. Each router R_d takes f as input and produces a scalar routing logit through a two-layer MLP with a ReLU activation. When incrementally training on domain d, all routers $\{R_{d'}\}_{d'=1}^{d}$ up to and including the current domain are set as trainable, while routers for future domains remain frozen.

Given a forward pass up to the current training domain d, the model iterates over all domains $d' \in \{1, \dots, d\}$, computing domain-specific features through the corresponding DS-BN pathway and producing a classifier output $\mathbf{c}_{d'} \in \mathbb{R}^K$ and a router logit $r_{d'} \in \mathbb{R}$ for each domain. All outputs are stacked and returned jointly as $\mathbf{C} \in \mathbb{R}^{d \times K}$ and $\mathbf{r} \in \mathbb{R}^d$, enabling the entire routing and classification decision to be performed within a single forward pass.

2.2. Incremental Training

When incrementally training on domain d, all model parameters are frozen except for the DS-BN layers, domain-specific classifier, and router module corresponding to domain d, as well as all router modules of previously learned domains. This selective unfreezing ensures that the convolutional weights and the previously learned classifiers remain intact, structurally preventing catastrophic forgetting of prior domain knowledge.

Router Label Assignment. The core of our training procedure is a novel router label assignment strategy. For each training sample, the router-weighted combination of classifiers is computed for all depths $d' \in \{1, \dots, d\}$:

$$\hat{\mathbf{c}}_{d'} = \frac{1}{d'} \sum_{t=1}^{d'} \alpha_t^{d'} \cdot \mathbf{c}_t, \quad \alpha_t^{d'} = \text{softmax}(\mathbf{r}_{1:d'})[t] \tag{1}$$

We then sequentially examine depths d' from 1 to $d-1$ to determine whether the weighted combination up to depth d' is sufficient to produce the correct prediction. When $\arg\max(\hat{\mathbf{c}}_{d'}) = y$, where $y \in \{1, \dots, K\}$ is the ground-truth class label, the sample is assigned a router label of d'. This implies sufficient knowledge up to domain d' for correct classification, thereby terminating the search. If no depth d' from 1 to $d-1$ yields the correct prediction, the sample is assigned to the current training domain d. This procedure encourages the router to maximize reuse of previously acquired knowledge while directing genuinely novel samples toward the newly introduced domain classifier.

Entropy Threshold. To further regulate the confidence of knowledge reuse, we optionally apply an entropy threshold τ_D during router label assignment. Even when $\arg\max(\hat{\mathbf{c}}_{d'}) = y$, the assignment to domain d' is accepted only if the prediction entropy falls below the threshold:

$$H(\hat{\mathbf{c}}_{d'}) = -\sum_{k=1}^{K} \hat{p}_k \log \hat{p}_k, \quad \hat{p}_k = \text{softmax}(\hat{\mathbf{c}}_{d'})[k] \tag{2}$$

If the entropy at domain d', denoted as $H(\hat{\mathbf{c}}_{d'})$, exceeds τ_D, the sample is not assigned to domain d' even if the prediction is correct, and the search continues to the next domain depth. This mechanism prevents the router from over-relying on uncertain prior-domain predictions, encouraging more conservative knowledge reuse under ambiguous conditions.

Training Objective. The total training loss is the sum of a classification loss $\mathcal{L}_{\text{class}}$ and a router supervision loss $\mathcal{L}_{\text{router}}$, both computed via cross-entropy:

$$\mathcal{L}_{\text{class}} = \text{CE}(\hat{\mathbf{c}}_{d^*}, y) \tag{3}$$

$$\mathcal{L}_{\text{router}} = \text{CE}(\mathbf{r}, d^*) \tag{4}$$

$$\mathcal{L} = \mathcal{L}_{\text{class}} + \mathcal{L}_{\text{router}} \tag{5}$$

where $y \in \{1, \dots, K\}$ is the ground-truth class label, $d^* \in \{1, \dots, d\}$ is the router label determined by the router label assignment procedure, and $\text{CE}(\cdot, \cdot)$ denotes the cross-entropy loss.

2.3. Inference

At inference time, domain identity is unavailable. Therefore, the input is sequentially evaluated by all domain branches learned up to the current incremental stage. This process produces classifier outputs $\mathbf{C} \in \mathbb{R}^{d \times K}$ and router logits $\mathbf{r} \in \mathbb{R}^d$, where d denotes the number of learned domains and K is the number of target classes.

$$d^* = \arg\min_{d'} H(\hat{\mathbf{c}}_{d'}), \quad d' \in \{1, \dots, d\} \tag{6}$$

The final class prediction is the argmax of softmax($\hat{\mathbf{c}}_{d^*}$). Unlike training, the entropy threshold is not applied during inference. Instead, the model computes predictions for all candidate domain depths and selects the aggregated output with the minimum prediction entropy, allowing the model to automatically select the most confident combination of accumulated domain knowledge without relying on any predefined confidence threshold.

3. EXPERIMENTAL SETUP

The audio preprocessing follows the same configuration as the baseline system. Audio clips are resampled to 32 kHz if necessary, then segmented into 4-second clips. Each clip is represented using log mel-band energies in 64 bands, with lower and upper cut-off frequencies of 50 Hz and 14 kHz respectively, using a Hamming window of 1024 samples with a hop size of 320 samples.

The model is optimized using the AMSGrad variant of the Adam optimizer with a learning rate of 1×10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$, and no weight decay. A cosine annealing learning rate scheduler is applied with a minimum learning rate of 1×10^{-3}. Incremental training is conducted for 1,000 epochs for both D2 and D3. No data augmentation is applied during training.

4. EXPERIMENTS

Table 1 presents the classification accuracy on the development test sets after sequentially learning Domains 2 and 3. For fair comparison, we reproduced the baseline system provided by the challenge organizers using the official pretrained checkpoint and evaluation protocol. Minor numerical differences from the values reported in the challenge description were observed, and all comparisons in this report are therefore based on our reproduced baseline results.

The proposed Router-based Progressive Knowledge Reuse framework consistently outperformed the baseline across all experimental settings. Even without applying the entropy-threshold strategy, the proposed method achieved an average accuracy of 59.06%, improving upon the baseline average accuracy of 53.47% by 5.59 percentage points. This result indicates that the introduction of domain-specific classifiers and router-guided knowledge aggregation alone provides substantial benefits for domain-agnostic incremental learning.

Applying the entropy-threshold strategy further improved performance. Among the evaluated thresholds, the most restrictive setting (ET = 0.01) achieved the best overall performance, yielding 70.03% accuracy on Domain 2 and 51.37% accuracy on Domain 3 after learning Domain 3. This corresponds to an average accuracy of 60.7%, representing a gain of 7.23 percentage points over the baseline and 1.64 percentage points over the variant without entropy thresholding.

Furthermore, the experimental results reveal that smaller entropy thresholds consistently produce better performance. As the threshold decreases from 0.1 to 0.01, the average accuracy increases from 58.99% to 60.7%. This suggests that conservative knowledge reuse—reusing prior domain knowledge only when the model is highly confident—is more beneficial than aggressive reuse under uncertain conditions. When the confidence of prior domain classifiers is insufficient, allowing the router to incorporate newly introduced domain-specific knowledge appears to result in more robust classification performance.

5. CONCLUSION

This paper presented a Router-based Progressive Knowledge Reuse framework for domain-agnostic incremental sound classification in DCASE 2026 Task 7. The proposed approach extends the DS-BN-based baseline by introducing domain-specific classifiers and lightweight routing modules that progressively determine how much previously acquired knowledge should be reused for each input sample. An entropy-threshold strategy was further incorporated during training to regulate the balance between knowledge reuse and knowledge expansion.

Experimental results on the development set demonstrated that the proposed method consistently outperformed the reproduced baseline across all evaluated settings. The best-performing configuration achieved 72.14% accuracy after learning Domain 2 and 60.7% average accuracy after learning Domain 3, corresponding to improvements of 12.72 and 7.23 percentage points over the baseline, respectively. The results indicate that progressively reusing accumulated domain knowledge through router-guided aggregation is an effective strategy for mitigating forgetting while improving domain-agnostic classification performance.

6. REFERENCES

[1] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

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

[3] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "Panns: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.

[4] M. Mulimani and A. Mesaros, "Domain-incremental learning for audio classification," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[5] ——, "Online incremental learning for audio classification using a pretrained audio model," in *2025 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2025, pp. 1–5.