Proteus with Isometric Domain Shaping: Domain-Agnostic Audio Classification for DCASE 2026 Task 7

Anonymous Submission
DCASE 2026 Task 7

***Abstract*—We present Proteus, a domain-agnostic audio classification system that addresses catastrophic forgetting in domain-incremental learning through Isometric Domain Shaping (IDS). The approach combines selective Low-Rank Adaptation (LoRA) with dual geometric regularization—Gram matrix isometry preservation and SIGReg feature space shaping—to enable sequential domain learning while maintaining knowledge from previously seen acoustic environments.**

Our architecture builds upon the CNN14 backbone from PANNs, applying LoRA adapters exclusively to the semantic layers (blocks 4-6) while keeping the frozen classification head and shared batch normalization statistics from the source domain (D1). The Isometric Domain Shaping framework enforces two complementary geometric constraints: (1) Gram Isometry Loss preserves relative pairwise feature geometry between a frozen teacher and the adapting student via L2-normalized Gram matrix alignment, and (2) SIGReg Loss shapes the feature manifold toward a zero-mean, isotropic Gaussian distribution by decorrelating dimensions and uniformizing variance.

The system achieves parameter efficiency by training only 0.13% of the full model (approximately 100K out of 75.6M parameters) through LoRA adapters with rank-8 decomposition and $\alpha = 16$ scaling. The teacher network is progressively updated after each domain to accumulate geometric knowledge across the sequence. Domain-agnostic inference is enabled through entropy-based task-head selection across seen domains, eliminating the requirement for domain identifiers at test time.

Experimental results on DCASE 2026 Task 7 demonstrate that Proteus with Isometric Domain Shaping maintains competitive accuracy across sequential acoustic domains while mitigating catastrophic forgetting through geometric manifold constraints rather than replay-based memorization.

***Index Terms*—domain-incremental learning, audio classification, continual learning, low-rank adaptation, manifold preservation, isometric regularization**

I. System Description

A. Overall Architecture

The **Proteus** system (named after the shape-shifting sea god) is a domain-agnostic CNN14-based architecture assembled entirely from modular components. The processing pipeline follows:

1) **Audio Input:** Raw waveform (32 kHz, 10-second clips)
2) **LogMelFrontend:** Spectrogram extraction with 64 mel bins
3) **Shared BatchNorm:** Over mel bands (frozen D1 statistics)
4) **ConvBlock 1-3:** Frozen backbone (1→64→128→256 channels)
5) **ConvBlock 4-6 with LoRA:** Trainable adapters (256→512→1024→2048 channels)
6) **Global Average Pooling**
7) **Frozen Linear Classifier:** 2048 → 10 classes

Key Design Principles:

- **Selective Adaptation:** Only the final three convolutional blocks (4-6) receive LoRA adapters, as these semantic layers are most sensitive to domain shift. Early layers (1-3) remain fully frozen to preserve low-level acoustic features.
- **Frozen Components:** The classification head and all batch normalization statistics are frozen at D1 values. This prevents the model from "drifting" into new decision boundaries and maintains alignment with the original feature space geometry.
- **Domain-Agnostic Design:** Unlike multi-head approaches, Proteus uses a single unified pathway with shared batch normalization, enabling inference without domain identification.

B. Isometric Domain Shaping (IDS)

Isometric Domain Shaping is the core innovation that enables continual learning without catastrophic forgetting. It enforces geometric constraints on the feature manifold through two complementary losses.

1) Gram Isometry Loss (Manifold Preservation): The Gram Isometry Loss preserves the **relative geometric structure** of learned feature representations during domain adaptation.

Formulation: Given student features $\mathbf{F}_s \in \mathbb{R}^{B\times D}$ and teacher features $\mathbf{F}_t \in \mathbb{R}^{B\times D}$ for a batch of size B with feature dimension D:

$$\mathbf{z}_s = \frac{\mathbf{F}_s}{\|\mathbf{F}_s\|_2} \quad \text{(L2 normalization)} \tag{1}$$

$$\mathbf{z}_t = \frac{\mathbf{F}_t}{\|\mathbf{F}_t\|_2} \quad \text{(L2 normalization)} \tag{2}$$

$$\mathbf{G}_{\text{student}} = \mathbf{z}_s \mathbf{z}_s^\top \quad \text{(pairwise cosine similarity)} \tag{3}$$

$$\mathbf{G}_{\text{teacher}} = \text{detach}(\mathbf{z}_t \mathbf{z}_t^\top) \quad \text{(no gradient)} \tag{4}$$

$$\mathcal{L}_{\text{gram}} = \|\mathbf{G}_{\text{student}} - \mathbf{G}_{\text{teacher}}\|_F^2 \quad \text{(Frobenius norm)} \tag{5}$$

Intuition: The Gram matrix captures pairwise relationships between samples in the feature space. By aligning the student's Gram matrix with the frozen teacher's, we preserve

relative distances and angles between feature embeddings—the "shape" of the learned manifold—rather than forcing absolute feature matching. This allows the model to adapt its representations to new domains while maintaining the discriminative geometry learned from previous domains.

Properties:

- **Translation Invariant:** Works even if feature centroids shift between domains
- **Scale Normalized:** L2 normalization focuses on angular relationships
- **Progressive Knowledge:** Teacher is updated after each domain to accumulate multi-domain geometry

2) SIGReg Loss (Feature Space Shaping): The SIGReg Loss shapes the feature manifold toward an **isotropic Gaussian distribution**, preventing feature collapse and promoting dimensional independence.

Formulation: Given features $\mathbf{F} \in \mathbb{R}^{B \times D}$:

$$\boldsymbol{\mu} = \text{mean}(\mathbf{F}, \text{dim} = 0) \tag{6}$$

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{F} - \boldsymbol{\mu}) \quad \text{(batch covariance)} \tag{7}$$

$$\begin{aligned} \mathcal{L}_{\text{sigreg}} = \lambda_{\text{offdiag}} \cdot \|\text{offdiag}(\boldsymbol{\Sigma})\|_F^2 \\ + \lambda_{\text{var}} \cdot \text{var}(\text{diag}(\boldsymbol{\Sigma})) \\ + \lambda_{\text{mean}} \cdot \|\boldsymbol{\mu}\|_2^2 \end{aligned} \tag{8}$$

Components:

- **Decorrelation (λ_{offdiag}):** Penalizes correlation between feature dimensions, ensuring each dimension captures independent information
- **Variance Uniformity (λ_{var}):** Ensures all dimensions contribute equally, preventing some dimensions from dominating
- **Zero Mean (λ_{mean}):** Centers the feature distribution, providing a stable geometric anchor

Benefits:

- Prevents feature space collapse during adaptation
- Maintains feature diversity across domains
- Complements Gram Isometry by enforcing intra-batch structure

C. Training Strategy

1) Selective LoRA Configuration: **Low-Rank Adaptation (LoRA)** injects trainable low-rank matrices into the frozen convolutional blocks. For each convolutional layer in blocks 4, 5, 6:

$$\mathbf{W}_{\text{conv}} = \mathbf{W}_{\text{frozen}} + \mathbf{AB} \cdot \frac{\alpha}{r} \tag{9}$$

where:

- $\mathbf{W}_{\text{frozen}}$: frozen pretrained weights
- $\mathbf{A} \in \mathbb{R}^{r \times k}$, $\mathbf{B} \in \mathbb{R}^{m \times r}$: trainable LoRA matrices
- $r = 8$: rank (controls capacity vs. efficiency tradeoff)
- $\alpha = 16.0$: scaling factor (preserves gradient magnitudes)
- $\mathbf{B}$ initialized to zeros (identity initialization at start)

Parameter Efficiency:

- Total parameters: 75.6M
- Trainable parameters (LoRA only): $\sim$100K (0.13%)
- Blocks adapted: 4, 5, 6 (semantic layers)
- Blocks frozen: 1, 2, 3 (low-level feature extractors)

2) Combined Loss Function: The total training loss for domain $\mathcal{D}_t$ combines cross-entropy with geometric regularizers:

$$\begin{aligned} \mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}}(\text{logits}, \text{labels}) \\ + \lambda_{\text{gram}} \cdot \mathcal{L}_{\text{gram}}(\mathbf{F}_s, \mathbf{F}_t) \\ + \lambda_{\text{sigreg}} \cdot \mathcal{L}_{\text{sigreg}}(\mathbf{F}_s) \end{aligned} \tag{10}$$

Typical hyperparameters: $\lambda_{\text{gram}} \in [0.1, 1.0]$ (typical: 0.5); λ_{sigreg} configured via $(\lambda_{\text{offdiag}}, \lambda_{\text{var}}, \lambda_{\text{mean}})$.

Loss Evolution During Training:

- $\mathcal{L}_{\text{CE}}$: Drives adaptation to the current domain's distribution
- $\mathcal{L}_{\text{gram}}$: Decreases as student aligns with teacher geometry (0.0097 $\rightarrow$ 0.0045 typical)
- $\mathcal{L}_{\text{sigreg}}$: Maintains feature space health by preventing collapse

3) Teacher Update Strategy: The teacher network is **progressively updated** after each domain completes:

1) Train student on domain $\mathcal{D}_t$
2) $\theta_{\text{teacher}} \leftarrow \text{deepcopy}(\theta_{\text{student}})$
3) Set teacher to eval mode: teacher.eval()
4) Freeze teacher parameters: teacher.requires_grad(False)

Rationale: This accumulates geometric knowledge across all seen domains. Without progressive updates, the teacher would only encode D1 geometry, causing progressive drift. With updates, the teacher represents a **consolidated manifold** spanning all previously learned distributions.

4) Batch Normalization Strategy: **Critical Design Decision:** All shared batch normalization layers are kept in **eval mode** during training. The shared BN statistics are ported from D1 and frozen. In train mode, standard BN would recompute running statistics from each small DDP-sharded batch, badly mis-scaling the frozen backbone and inflating the loss. Forcing eval mode keeps the model using D1 running statistics, which empirically converges $\sim$2$\times$ faster and achieves higher accuracy (e.g., 0.65 vs 0.60 on D2 at 4 epochs).

D. Inference Strategy

Since the system uses **shared batch normalization**, there is only one forward pass needed per sample. The model naturally produces domain-agnostic predictions without requiring domain identification.

For multi-head variants, entropy-based task selection is used:

1) For each seen task head $t \in \{0, \ldots, t_{\text{current}}\}$:
   - Compute probabilities: $\mathbf{p}_t \leftarrow \text{softmax}(f(\mathbf{x}, t))$
   - Compute entropy: $H_t \leftarrow -\sum_c \mathbf{p}_t[c] \log \mathbf{p}_t[c]$
2) Select head with lowest entropy: $t^* \leftarrow \arg\min_t H_t$
3) Make prediction: $\hat{y} \leftarrow \arg\max_c f(\mathbf{x}, t^*)[c]$

Key Property: No domain identifier is required at test time. The model's learned geometry enables it to generalize across all seen acoustic environments through a unified decision boundary.

E. Evaluation Metrics

The system reports metrics aligned with DCASE 2026 Task 7 requirements:

- **Entropy-Based DIL Accuracy:** Average accuracy across all seen domains using entropy-based head selection
- **Per-Domain Micro Accuracy:** Plain sample accuracy per domain using the correct head
- **Per-Domain Macro Accuracy:** Mean of class-wise recalls over classes present in each domain (**competition ranking metric**)
- **Average Forgetting:** Average degradation from best prior accuracy to final accuracy
- **Sweep Objective:** Combined metric: $\text{macro}_{\text{overall}} - 0.5 \times$ avg_forgetting

II. Technical Implementation

A. Model Configuration

Architecture Specifications:

- **Backbone:** CNN14 (PANNs)
- **Input:** 32 kHz audio waveforms (10-second clips)
- **Frontend:** Log-Mel spectrogram (window: 1024, hop: 320, mel bins: 64, fmin: 50 Hz, fmax: 14 kHz)
- **Convolutional Blocks:** 6 blocks (1→64→128→256→512→1024→2048 channels)
- **Pooling:** Global Average Pooling (GAP)
- **Classifier:** Linear (2048 → 10 classes, frozen)

LoRA Configuration:

- Applied to: Blocks 4, 5, 6 only
- Rank: 8, Alpha: 16.0
- Initialization: $\mathbf{B} = 0$ (identity at start)
- Trainable params: ∼100K (0.13% of 75.6M)

B. Training Configuration

Optimizer: AdamW

- Learning rate (D1): 10^{-3} (initial phase)
- Learning rate (D2+): 10^{-4} (incremental phases)
- Scheduler: Cosine annealing with warmup (optional)

Regularization Hyperparameters:

- Gram Isometry: $\lambda_{\text{gram}} \in [0.1, 1.0]$ (typical: 0.5)
- SIGReg: $\lambda_{\text{offdiag}}, \lambda_{\text{var}} \in [0.1, 2.0]$, $\lambda_{\text{mean}} \in [0.01, 1.0]$
- Dropout: 0.2 (in LoRA blocks)

Training Protocol:

- Epochs per domain: 4-10 (configurable)
- Batch size: 32-64 (per GPU)
- Mixed Precision: AMP (fp16) enabled
- Gradient Clipping: max_norm=1.0
- DDP: Multi-GPU training supported

III. Conceptual Foundations

A. Why "Isometric Domain Shaping"?

The term **Isometric Domain Shaping** captures two key ideas:

1) **Isometric (Gram Isometry):** Preserves **relative distances and angles** in the feature manifold, ensuring that the geometric "shape" of learned representations remains consistent across domains. This is analogous to an isometric transformation in geometry—one that preserves distances and angles.
2) **Domain Shaping (SIGReg):** Actively **shapes** the feature distribution toward a canonical form (zero-mean, isotropic Gaussian), preventing domain-specific collapse and ensuring dimensional independence.

Together, these mechanisms enable the model to **adapt its form** (like the mythological Proteus) to each domain while maintaining a **coherent geometric structure** across the sequence.

B. Comparison to Related Approaches

TABLE I
Comparison with Continual Learning Methods

Method	Domain ID?	Replay?	Param Eff.
Proteus (IDS)	No	No	0.13%
EWC	No	No	All trainable
PackNet	No	No	Fixed capacity
iCaRL	No	Yes	All trainable
Multi-head	Yes	No	Grows linearly

Key Advantages:

- No replay buffer (privacy/storage)
- Parameter efficient (0.13% trainable)
- Domain-agnostic (no domain ID needed)
- Geometric preservation (manifold structure)

C. Theoretical Intuition: Why Gram Matrices?

The Gram matrix $\mathbf{G} = \mathbf{Z}\mathbf{Z}^\top$ (where $\mathbf{Z}$ is L2-normalized features) captures **all pairwise cosine similarities** in a batch:

$$\mathbf{G}[i, j] = \cos(\text{angle}(\mathbf{z}_i, \mathbf{z}_j)) = \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{z}_i\|\|\mathbf{z}_j\|} \tag{11}$$

Why this matters for continual learning:

1) **Manifold Topology:** The Gram matrix encodes the **local topology** of the feature manifold. By preserving $\mathbf{G}$, we preserve the relative arrangement of samples in feature space.
2) **Class Separation:** Classes that were well-separated in the teacher's feature space remain separated in the student's adapted space.
3) **Scale Invariance:** L2 normalization removes feature magnitude variations, focusing purely on **directional relationships**.
4) **Efficiency:** Computing $\mathbf{G}$ requires only a single batch-wise matrix multiplication, making it computationally cheap.

D. Complementarity of Gram Isometry and SIGReg

TABLE II
COMPLEMENTARY REGULARIZATION MECHANISMS

Aspect	Gram Isometry	SIGReg
Scope	Inter-sample	Intra-batch
Target	Teacher geometry	Canonical dist.
Invariance	Relative structure	Absolute structure
Purpose	Prevent forgetting	Prevent collapse

Synergy: Gram Isometry prevents the model from "forgetting" the relative arrangement learned from previous domains, while SIGReg prevents the adapted features from collapsing into degenerate subspaces. Together, they create a **stable, well-conditioned feature space** that can accommodate multiple domain distributions.

IV. EXPERIMENTAL RESULTS

A. DCASE 2026 Task 7 Performance

TABLE III
SYSTEM PERFORMANCE ON DCASE 2026 TASK 7

Metric	Value	Description
D2 Accuracy	76.44%	Micro accuracy on D2 after adaptation
D2+D3 Average	61.32%	Average micro accuracy after D3
Avg Forgetting	$\sim$0.15	Typical forgetting rate
Macro Overall	Competitive	Competition ranking metric

B. Key Observations

1) **Gram Loss Decay:** $\mathcal{L}_{\text{gram}}$ typically decreases from $\sim$0.0097 to $\sim$0.0045 during domain training, indicating successful geometric alignment.
2) **Parameter Efficiency:** Training only 100K parameters (0.13%) enables fast adaptation per domain ($<$ 1 hour on single GPU for 4 epochs).
3) **Domain-Agnostic Inference:** Entropy-based selection or direct single-head inference both achieve competitive accuracy without domain labels.
4) **Forgetting Mitigation:** Isometric Domain Shaping reduces but does not eliminate catastrophic forgetting. The geometric constraints provide a **soft anchor** to previous knowledge without preventing all adaptation.

V. CONCLUSION

The **Proteus** system with **Isometric Domain Shaping (IDS)** offers a parameter-efficient, replay-free approach to domain-incremental audio classification. By combining selective LoRA adaptation with dual geometric regularization (Gram Isometry + SIGReg), the system preserves learned feature manifolds while adapting to new acoustic domains. The domain-agnostic design eliminates the need for domain identification at test time, making it practical for real-world deployment where domain boundaries are unknown.

Key Contributions:

1) **Geometric Regularization Framework:** Dual-loss design (Gram + SIGReg) that preserves manifold topology and prevents feature collapse
2) **Selective Adaptation:** LoRA applied only to semantic layers (blocks 4-6), balancing adaptability and stability
3) **Progressive Teacher Updates:** Accumulates multi-domain knowledge across the sequence
4) **Domain-Agnostic Inference:** Single unified model without explicit domain identification
5) **Parameter Efficiency:** 0.13% trainable parameters enable fast, memory-efficient adaptation

Future Directions:

- Adaptive regularization based on forgetting/adaptation tradeoff
- Contrastive objectives for stronger inter-class separation
- Ensemble teachers for richer geometric constraints
- Automatic domain shift detection for adaptive training schedules

ACKNOWLEDGMENTS

[To be populated]