Beyond Euclidean Assumptions: Geometry-Aware Adaptive Routing for Remote Sensing Segmentation

Jie Qiu[1], **Dizuo Cao**[1], **Linwei Dai**[2], **Xin Li**[3], **Fan Yang**[3], **Dong Yu**[4], **Changying Wang**[1], **Zongheng Wen**[5], **Youqin Chen**[6*], **Jianzhang Chen**[1*]
[1]College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou, China
[2]iFLYTEK, Hefei, China
[3]AIQ, Abu Dhabi, UAE
[4]School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China
[5]School of Information Science and Technology, Xiamen University Tan Kah Kee College, Zhangzhou, China
[6]School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou, China
{jieqiu_fafu, caodizuo, wangchangying, jzchen}@fafu.edu.cn, 21171213925@stu.xidian.edu.cn,
{xinli_uestc, fanyang_uestc}@hotmail.com, 23115071@bjtu.edu.cn, zhwen@xujc.com, chenyouqin@fjut.edu.cn

Abstract

Remote sensing imagery poses a distinct challenge for semantic segmentation due to its inherent fractal complexity and the diversity of geometric structures present in real-world geospatial scenes. Euclidean-based models typically assume spatial uniformity; however, such assumptions often break down when confronted with objects exhibiting markedly different structural characteristics—such as roads versus vegetation—thereby complicating the feature representation process. Hyperbolic space offers a theoretically grounded alternative for modeling such hierarchical and heterogeneous patterns, yet fully replacing Euclidean geometry incurs significant computational overhead. We therefore introduce Geometry-Aware Adaptive Routing (GAAR), a novel module that facilitates geometry-aware routing by dynamically allocating high-level features to either Euclidean or Hyperbolic subspaces through a learnable binary gating mechanism, informed by structural priors learned during training. To further promote routing stability and geometric consistency, we introduce Geometry-Aware Deterministic Regularization (GADR), a regularization strategy that encourages confident, structure-aligned assignments. GAAR is plug-and-play and integrates seamlessly into existing segmentation architectures. Experiments on three challenging Remote Sensing Image Semantic Segmentation (RSISS) benchmarks demonstrate that our approach consistently outperforms state-of-the-art (SOTA) methods, particularly in geometrically complex regions, offering a scalable and effective solution to the limitations of purely Euclidean modeling.

Code — https://github.com/HeryJie/GAAR.

Introduction

Remote Sensing Image Semantic Segmentation (RSISS) stands as a cornerstone task in large-scale scene understanding, with far-reaching applications in land cover analysis (Li et al. 2022), geospatial object recognition (Griffiths and Boehm 2019), precision agriculture (Luo et al. 2024), and environmental monitoring (Samie et al. 2020). The central objective is to predict a semantic label for every pixel in high-resolution overhead imagery, facilitating structured interpretation of complex geographic environments.

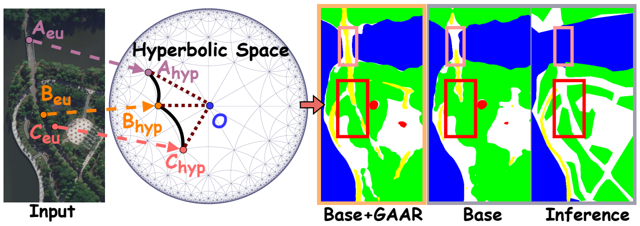

Figure 1: Feature distribution and segmentation results (Euclidean *vs.* Hyperbolic space). In the baseline model (middle), road features (A_{eu}, B_{eu}) and nearby trees (C_{eu}) are entangled due to structural complexity and fractal dimension (FD) heterogeneity, resulting in blurred boundaries. GAAR introduces Hyperbolic embeddings (left), which tightly cluster road features (A_{hyp}, B_{hyp}) and clearly separate high-FD vegetation (C_{hyp}), enhancing both semantic discrimination and topological continuity in the final segmentation (right).

Modern segmentation pipelines typically rely on a Euclidean geometric assumption, where spatial continuity and local context guide the learning of discriminative representations. While effective for low-level patterns, this assumption breaks down in structurally heterogeneous scenes, especially in geospatial settings where objects follow different organizational principles. For example, roads and vegetation differ not only in appearance but also in fractal dimension (FD), which reflects spatial complexity. These differences give rise to nonlinear, hierarchical, and scale-varying geometric structures. Euclidean space, with its flat and uniform metric, often fails to encode such variations, leading to feature crowding, semantic ambiguity, and topological distortion. This limitation, which we refer to as the Euclidean representational bottleneck, frequently results in over-smoothing or misclassification in geometrically complex regions (Peng

---

[*] Corresponding authors.

et al. 2022). From this perspective, compared with flat Euclidean space, high-FD structures can be represented more faithfully in Hyperbolic manifolds that possess stronger negative curvature.

To overcome these limitations, Hyperbolic geometry has recently emerged as a promising alternative. With its negative curvature and exponential expansion, Hyperbolic space naturally accommodates hierarchical and self-similar structures (Ganea, Bécigneul, and Hofmann 2018; Liu, He, and Han 2025), offering superior separation in feature space for semantically distant yet visually adjacent objects. However, naively replacing Euclidean geometry with Hyperbolic formulations across the entire network introduces significant computational and optimization burdens (Chen et al. 2024). These observations motivate two core questions. *Question 1*: How can Hyperbolic representations be embedded into deep networks in a manner that is both computationally efficient and aligned with task-specific requirements? *Question 2*: Can we design a unified, data-driven mechanism that dynamically determines, for each feature, whether Euclidean or Hyperbolic space offers a more suitable representational geometry, based on the intrinsic structural complexity of the input? Addressing these questions is key to unlocking the full potential of hybrid geometric reasoning in deep visual understanding.

In response to these challenges, we introduce Geometry-Aware Adaptive Routing (GAAR), a lightweight and modular architectural unit that dynamically routes high-level features to either Euclidean or Hyperbolic space via a differentiable, structure-aware gating mechanism. This mechanism is parameterized through a Gumbel-Softmax reparameterization strategy, enabling soft and learnable space selection that evolves into deterministic routing via temperature annealing. To enhance the confidence and geometric consistency of routing decisions, we further introduce Geometry-Aware Deterministic Regularization (GADR), a geometry-aware regularization term that encourages the model to converge toward bimodal, structure-aligned space assignments, thereby promoting clearer separation between complex and regular feature patterns.

As illustrated in Figure 1, GAAR enables geometry-aware representation by dynamically routing high-level features to either Euclidean or Hyperbolic space based on structural complexity. In the baseline model, roads (e.g., A_{eu}, B_{eu}) and adjacent vegetation (C_{eu}) are embedded in Euclidean space, where differences in FD are not adequately captured, leading to entangled features and degraded topological coherence, especially when road segments are occluded. With GAAR, geometry-sensitive features are routed to Hyperbolic space, resulting in improved separability between semantically distinct regions (e.g., B_{hyp} vs. C_{hyp}) and tighter clustering of same-class structures (e.g., A_{hyp} and B_{hyp}), thereby enhancing semantic continuity. Critically, GAAR achieves this without imposing architectural constraints: it is fully differentiable, plug-and-play, and computationally lightweight, supporting seamless end-to-end integration with limited overhead.

Experimental results across three challenging remote sensing benchmarks demonstrate that GAAR can be seamlessly integrated into a variety of SOTA segmentation architectures, consistently addressing the representational limitations of Euclidean-based modeling. Our approach achieves substantial gains in key evaluation metrics, including mean Intersection over Union (mIoU), mean F1-score, and Overall Accuracy (OA), surpassing competitive baselines, particularly in geometrically complex regions. Our main **contributions** are as follows:

- We propose **GAAR**, a lightweight and fully differentiable module that enables per-channel routing between Euclidean and Hyperbolic spaces. Guided by learned structural priors, GAAR enhances the network's capacity to model hierarchical and fractal patterns while introducing low computational overhead.
- We introduce **GADR**, a geometry-guided regularization term that encourages confident and structure-aware space selection. Leveraging Gumbel-Softmax reparameterization and temperature annealing, GADR enables near-deterministic, interpretable routing in a fully differentiable manner.
- We demonstrate that our method consistently achieves **SOTA performance** across multiple RSISS benchmarks and backbone architectures, delivering robust generalization and improved geometric fidelity while remaining compatible with standard training pipelines and deployment environments.

Related Work

Remote Sensing Image Semantic Segmentation Remote Sensing Image Semantic Segmentation (RSISS) has become a central task in geospatial vision, driven by advances in both convolutional and transformer-based architectures. Early efforts such as Fully Convolutional Networks (FCNs) (Long, Shelhamer, and Darrell 2015) demonstrated the viability of applying Convolutional Neural Networks (CNNs) (LeCun, Bengio, and Hinton 2015) to dense prediction tasks in remote sensing. Building on this foundation, subsequent models introduced progressively stronger spatial reasoning capabilities: ResNet (He et al. 2016) enabled deeper feature extraction via residual learning, DeepLab (Chen et al. 2018b,a) captured multi-scale context through atrous spatial pyramid pooling, and DANet (Fu et al. 2019) incorporated attention mechanisms to enhance global feature integration. More recent designs, such as LSKNet (Li et al. 2024) and UNetFormer (Wang et al. 2022b), refined encoder-decoder architectures with scale-aware kernels and lightweight attention modules, while In2NeCT (Shen et al. 2025) focused on optimizing intra- and inter-class feature separability. Beyond CNN-based and point-based refinement methods such as PointRend (Kirillov et al. 2020), transformer-based models have gained prominence for their ability to model long-range dependencies in RSISS. Segmenter (Strudel et al. 2021) and DC-Swin (Wang et al. 2022a) leverage ViT (Dosovitskiy et al. 2021) and Swin Transformer (Liu et al. 2021) backbones to encode rich global context, while GSNet (Ye, Zhuge, and Zhang 2025) integrates CLIP (Radford et al. 2021) to unify vision-language priors with remote sensing semantics.

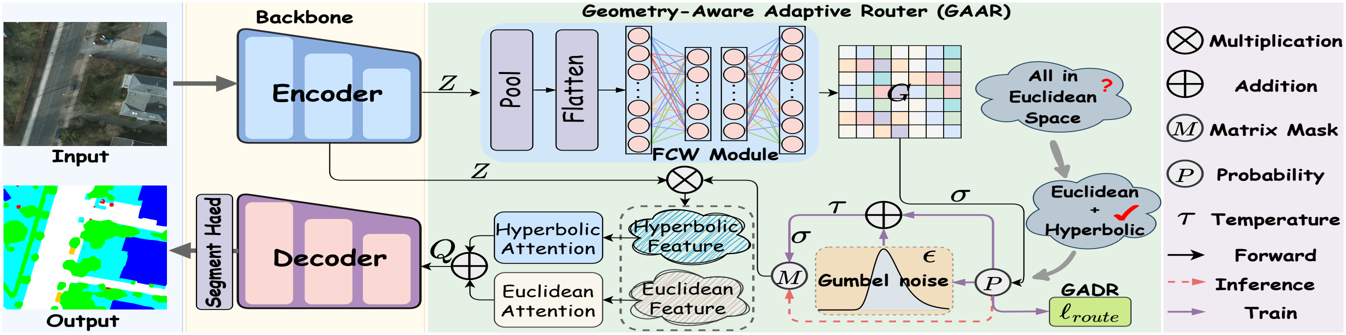

Figure 2: Overview of the proposed framework. GAAR adaptively routes feature channels to Euclidean or Hyperbolic branches based on learned probabilities, with GADR regularization promoting confident, geometry-aware assignments. The fused output is decoded to produce the final segmentation map.

Despite existing progress, most methods still implicitly employ Euclidean geometry to model spatial relationships, which works well in regular or low-complexity scenarios but falls short when dealing with geospatial structures exhibiting hierarchical, nonlinear, or fractal features. Particularly when objects with significantly different fractal dimensions—such as roads and vegetation—coexist, this often leads to feature entanglement and inaccurate boundary localization. Addressing this geometric mismatch is critical for improving the robustness and structural awareness of remote sensing segmentation.

Hyperbolic Space Hyperbolic geometry has recently emerged as a compelling alternative to Euclidean space for modeling hierarchical and non-Euclidean structures in computer vision. Owing to its negative curvature and exponential expansion, Hyperbolic space is particularly well-suited for representing data with latent tree-like or fractal organization (Khrulkov et al. 2020; Yan et al. 2021). Ganea et al. (Ganea, Bécigneul, and Hofmann 2018) laid the theoretical groundwork by formulating Hyperbolic analogues of core deep learning components, including embedding layers and activation functions, thereby enabling the integration of Hyperbolic geometry into end-to-end neural architectures. Building upon this foundation, Ghadimi Atigh et al. (GhadimiAtigh et al. 2022) extended Hyperbolic modeling to dense prediction by proposing a pixel-wise classification framework in Hyperbolic space, demonstrating its feasibility for segmentation tasks. More recently, Chen et al. (Chen et al. 2024) leveraged Hyperbolic distance metrics to design a geometry-aware uncertainty loss that enhances boundary discrimination in regions of high prediction ambiguity.

Despite significant progress, previous studies have primarily focused on fully replacing Euclidean geometry with Hyperbolic geometry, which, although effective in capturing hierarchical semantics, often incurs high computational costs and limits model flexibility. In contrast, we propose a hybrid geometric paradigm that adaptively assigns features to either Euclidean or Hyperbolic space based on their complexity, leveraging the complementary strengths of both geometries to more efficiently and expressively model the heterogeneous characteristics of remote sensing imagery.

Methodology

We propose a novel geometry-aware segmentation framework composed of two key components: GAAR and GADR. As illustrated in Figure 2, the overall architecture consists of three modular elements: a RSISS backbone, the GAAR module for dynamic geometry routing, and the GADR term for enforcing consistent, structure-aligned assignments.

The framework is architecture-agnostic and introduces no constraints on backbone design. GAAR is fully differentiable and plug-and-play, enabling seamless integration into both CNN and Transformer-based models with minimal modification. This design supports efficient, geometry-aware representation learning within diverse RSISS pipelines. The same geometry-aware, sparsely mixed Euclidean–Hyperbolic design may also serve as a generic plug-in for structurally complex vision tasks, including whole-slide image classification (Zhao et al. 2025a), universal and semi-supervised medical image segmentation (Zhao et al. 2025b; Zhang et al. 2023), and point cloud semantic segmentation (Yang, Li, and Shen 2022).

Overall Architecture

The overall pipeline is illustrated in Figure 2. Formally, given an image $I \in \mathbf{R}^{H\times W\times 3}$, a feature encoder derived from from any standard RSISS architecture first extracts high-level features $Z \in \mathbf{R}^{H\times W\times C}$.

These features are then passed to the proposed GAAR module, which performs per-channel geometry-aware routing based on learned structural priors. For each channel, GAAR dynamically selects the more suitable embedding space, *i.e.,* Euclidean or Hyperbolic, depending on the underlying spatial complexity. Features routed to the Hyperbolic branch are projected onto the Poincaré manifold, while those retained in Euclidean space remain in their original representation. The two branches are subsequently fused to obtain the unified representation $Q \in \mathbf{R}^{H\times W\times C}$.

Finally, Q is passed through the decoder and segmentation head to produce the final semantic prediction $O \in \mathbf{R}^{H\times W\times N}$, where N denotes the number of semantic categories. This design preserves architectural compatibility while enhancing geometric expressiveness, enabling more effective segmentation in structurally diverse regions.

Geometry-Aware Adaptive Routing Module

Probabilistic Modeling of High-Level Features High-level features in remote sensing images are inherently abstract and structurally diverse, often exhibiting varying degrees of geometric complexity. To address *Question 1* posed in the Introduction—namely, how to selectively incorporate Hyperbolic modeling while maintaining computational efficiency—we formulate geometry selection as a channel-wise probabilistic routing task.

Given the feature map $Z \in \mathbf{R}^{B\times C\times H\times W}$ extracted from the encoder, we introduce a lightweight Feature Channel Weight (FCW) module that aggregates spatial information and produces a set of channel-wise logits $G \in \mathbf{R}^{B\times C}$. These logits are passed through a sigmoid activation function $\sigma(\cdot)$ to produce a probabilistic mask $P \in \mathbf{R}^{B\times C}$:

$$P = \sigma(G), \tag{1}$$

where each element P_i represents the probability of assigning the i-th feature channel to the Hyperbolic space (with $1 - P_i$ corresponding to Euclidean space).

The resulting soft mask $M = P$ guides feature routing into the appropriate geometric branch. During training, the FCW module is optimized via backpropagation to learn structure-aware probabilities that reflect the latent geometry of each channel. Importantly, this formulation enables a differentiable approximation of discrete space selection, allowing the model to softly allocate features while preserving end-to-end trainability.

By modeling routing decisions in a probabilistic and geometry-aware manner, the network learns to adaptively balance representational expressiveness and computational cost, thereby addressing the fundamental trade-off between geometric precision and efficiency.

Generation of the Soft Mask Matrix M To enable differentiable routing across geometric branches, we construct the binary mask M via a soft approximation using the Gumbel-Softmax reparameterization (Dang et al. 2022). Instead of applying non-differentiable thresholding to the routing probabilities $P \in \mathbf{R}^{B\times C}$, we simulate discrete decisions by injecting stochastic noise into the log-probability space.

Concretely, we sample i.i.d. uniform random variables $U \sim \text{Uniform}(0, 1)$ with the same shape as P and compute Gumbel perturbations as:

$$\epsilon_i = -\log(-\log(u_i)), \quad u_i \in U. \tag{2}$$

The perturbed logits are obtained as $\log(P_i + \omega) + \epsilon_i$, where $\omega = 10^{-8}$ is a small constant to ensure numerical stability. We scale these by a temperature parameter τ and apply the sigmoid function $\sigma(\cdot)$ to yield the soft mask:

$$M_i = \sigma\left(\frac{\log(P_i + \omega) + \epsilon_i}{\tau}\right). \tag{3}$$

Unlike softmax-based sampling, the sigmoid function allows for independent routing across channels. Each M_i represents the probability of assigning the i-th channel to the Hyperbolic space, while $1-M_i$ corresponds to the Euclidean branch. The resulting mask is differentiable, enabling end-to-end optimization of routing behavior via backpropagation.

Dynamic Temperature Annealing The temperature parameter τ governs the sharpness of the soft routing distribution produced by the Gumbel-Softmax relaxation. At the beginning of training, a high temperature encourages near-uniform values in M, facilitating stochasticity and exploration across possible geometry assignments. As training proceeds, τ is progressively annealed according to an exponential decay schedule:

$$\tau^{(t)} = \max\left(\tau^{(t-1)} \cdot e^{-\beta t}, \tau_{\min}\right), \tag{4}$$

where $\beta = 3 \times 10^{-5}$ is the decay rate, $\tau^{(0)} = 1.0$, and $\tau_{\min} = 0.1$ is the minimum temperature threshold.

This annealing process gradually sharpens the distribution of M, pushing its values toward binary extremes and promoting more confident, geometry-aware routing decisions. During inference, the soft probabilities P are binarized via deterministic thresholding, yielding discrete and stable channel assignments.

By combining differentiable sampling in training with hard decision-making at inference, this hybrid mechanism supports both gradient-based learning and efficient deployment, effectively bridging the gap between optimization flexibility and execution determinism.

Dual-Stream Space Selection The soft mask $M \in \mathbf{R}^{B\times C}$ is used to dynamically partition the feature channels of $Z \in \mathbf{R}^{B\times C\times H\times W}$ into two geometric streams. Channels with low routing probabilities in M are assigned to the Euclidean branch, where they are processed via a standard multi-head self-attention mechanism (Vaswani et al. 2017), denoted as Euclidean Attention. The output of this branch is written as `euc_out`.

Conversely, channels with high probabilities (values close to 1) are routed to the Hyperbolic branch. These features are first projected from the tangent space at the origin onto the Poincaré ball $\mathbf{B}_c^d$ via the exponential map $\exp_0^c(\cdot)$, and then passed through a Hyperbolic self-attention module (HDN), architecturally mirroring its Euclidean counterpart. The output is then mapped back to the tangent space using the logarithmic map $\log_0^c(\cdot)$, yielding the Hyperbolic output `hyp_out`. The two branches are combined to form the final feature map $Q \in \mathbf{R}^{B\times C\times H\times W}$:

$$\texttt{hyp\_out} = \log_0^c\left(\text{HDN}\left(\exp_0^c(Z \cdot M)\right)\right), \tag{5}$$

$$Q = \texttt{euc\_out} + \texttt{hyp\_out}. \tag{6}$$

Here, the exponential map projects a vector v from the tangent space onto the Poincaré ball:

$$\exp_0^c(v) = \tanh\left(\sqrt{c}\|v\|\right) \cdot \frac{v}{\sqrt{c}\|v\|}, \tag{7}$$

and the logarithmic map performs the inverse projection:

$$\log_0^c(y) = \frac{1}{\sqrt{c}} \tanh^{-1}\left(\sqrt{c}\|y\|\right) \cdot \frac{y}{\|y\|}. \tag{8}$$

The scalar $c > 0$ is the curvature parameter that controls the degree of negative curvature in the Hyperbolic space. A higher value of c corresponds to stronger Hyperbolicity, allowing the model to better separate features with hierarchical or fractal complexity.

Geometry-Aware Deterministic Regularization

To encourage confident and deterministic routing during training, we introduce a regularization term that penalizes uncertainty in the learned routing probabilities. Specifically, we define the GADR loss as the expected binary variance of the routing scores $p_i \in P$, given by:

$$\mathcal{L}_{\text{route}} = \lambda \cdot \mathbf{E}_Z \left[\frac{1}{C} \sum_{i=1}^{C} p_i(1 - p_i) \right], \tag{9}$$

where C denotes the number of channels, λ is a weighting hyperparameter, and $\mathbf{E}_Z$ denotes the expectation over the input feature distribution. Intuitively, this regularizer reaches its maximum when $p_i = 0.5$, corresponding to maximal uncertainty, and vanishes as p_i approaches 0 or 1—thereby promoting confident, near-binary routing behavior.

By minimizing $\mathcal{L}_{\text{route}}$, the network is encouraged to make unambiguous assignments, pushing each feature channel decisively toward either the Euclidean or Hyperbolic branch. This yields sparser and more interpretable routing masks, which align well with the goal of lightweight and geometry-aware inference.

We incorporate the GADR regularization into the total training objective by combining it with the segmentation loss from the RSISS backbone (Wang et al. 2022b; Li et al. 2024) :

$$\mathcal{L} = \mathcal{L}_{\text{rsiss}} + \mathcal{L}_{\text{route}}. \tag{10}$$

Furthermore, the outputs from the Euclidean and Hyperbolic branches are fused via element-wise addition to form the final representation. This fusion enhances the complementarity between geometries and contributes to improved segmentation performance. In doing so, our framework addresses the challenge of adaptive geometric selection highlighted in *Question 2* of the Introduction.

Experiments

Datasets & Evaluation Metrics

We evaluate our method on three remote sensing semantic segmentation benchmarks: LoveDA (Wang et al. 2021), ISPRS Vaihingen, and ISPRS Potsdam (Rottensteiner et al. 2020). LoveDA provides 30 cm GSD images across seven categories (Bld, Rd, Wat, Bar, For, Agr, Bkg), with image IDs 0–2521 for training and 2522–4190 for testing. ISPRS Vaihingen contains 9 cm GSD images with five foreground classes and one background class, while Potsdam offers 5 cm GSD images with the same six categories. For both ISPRS datasets, we adopt the official UNetFormer split (Wang et al. 2022b). Performance is assessed using three standard metrics, *i.e.,* overall accuracy (OA), average F1, and mean IoU (mIoU), capturing pixel correctness, class-level balance, and region overlap quality.

Implementation Details

To demonstrate GAAR's generality, we integrate it into several SOTA segmentation models with diverse backbones: UNetFormer (Wang et al. 2022b) with ResNet18/34/50 (He et al. 2016), LSKNet (Li et al. 2024) with LSKNet-T/S, and DC-Swin (Wang et al. 2022a) based on the Swin Transformer (Liu et al. 2021) (Swin-T, Swin-S, Swin-B). This setup covers both CNN and Transformer families. All experiments are run in PyTorch on a single RTX 4090 using AdamW, keeping all other hyperparameters unchanged from the original implementations (Wang et al. 2022b; Li et al. 2023). Following (Wang et al. 2022b), we apply random scaling (0.75–1.5), flipping, rotation, and 512×512 cropping for training, and use multi-scale testing with test-time flipping during inference.

Main Results

Quantitative Evaluation Table 1 shows consistent gains after adding GAAR. On LoveDA, the best model reaches 55.49 mIoU. On ISPRS Potsdam and Vaihingen, GAAR achieves 93.54 / 91.70 F1, 92.59 / 92.07 OA, and 88.15 / 84.59 mIoU. Within ResNet–UNetFormer, mIoU increases by +1.82, +2.22, and +2.53 for ResNet-18/34/50, with the *Barren* class improving most. Transformer backbones show the same trend: on DC-Swin, Swin-S performs best, while Swin-B and Swin-T gain +1.21 and +1.24; in LSKNet, GAAR adds +1.93 (LSKNet-T) and +2.08 (LSKNet-S). Across 16 class-wise F1 scores, 10 rise by more than 0.3, with mIoU jumps of +0.96 on Potsdam (LSKNet-S) and +0.93 on Vaihingen (LSKNet-T), and OA gains up to +0.74 and +0.43. Overall, GAAR strengthens feature separation in geometrically complex or topologically ambiguous regions.

Qualitative Evaluation Figure 3 shows qualitative results before and after adding GAAR to UNetFormer on Potsdam, Vaihingen, and LoveDA. GAAR consistently improves segmentation in structurally complex or low-contrast regions. On Potsdam, it sharpens boundaries between the visually similar Building and Tree classes, which are often confused due to color proximity but separable by their differing fractal structures. On Vaihingen, GAAR yields more complete and contiguous Building masks, reducing the fragmentation seen in the baseline. On LoveDA, it produces more continuous Road predictions, especially in sparsely annotated or occluded areas. Across datasets, the highlighted regions show GAAR's ability to resolve semantic ambiguities and recover topology lost in purely Euclidean encoders.

Ablation Study

To evaluate the contribution of each component in GAAR, we conduct ablation studies on the LoveDA dataset using UNetFormer with ResNet18, ResNet34, and ResNet50 encoders. mIoU is used as the primary evaluation metric.

Component-Wise Analysis Table 2 reports ablations for the main components in GAAR. The FCW module is essential; removing it causes a complete performance collapse across all backbones, including a 7.25 mIoU drop on ResNet18, showing its importance for routing. Removing the Hyperbolic or Euclidean attention branch also results in notable declines (up to 2.09 and 1.85 mIoU on ResNet18 and ResNet50), confirming the complementarity of the dual-geometry design. We further evaluate three coupled mechanisms: GADR regularization, Gumbel-based stochastic routing, and temperature annealing. Removing any one leads to

Method		LoveDA								Vaihingen			Potsdam		
	Backbone	Bkg	Bld	Rd	Wat	Bar	For	Agr	mIoU	F1	OA	mIoU	F1	OA	mIoU
DeepLabV3+ (Chen et al. 2018b)	ResNet50	46.83	49.48	50.40	67.44	22.41	43.30	52.65	47.50	81.22	86.16	79.61	87.87	86.18	78.65
Segmenter (Strudel et al. 2021)	ViT-B	48.02	50.93	50.36	65.69	20.17	42.43	47.82	46.48	81.89	85.25	76.08	82.52	82.24	80.64
DANet (Fu et al. 2019)	ResNet50	53.67	59.62	51.89	66.01	29.34	43.33	53.25	51.02	85.86	89.23	75.95	90.41	89.82	82.65
MANet (Li et al. 2021)	ResNet50	54.65	62.35	52.64	67.73	24.31	44.28	53.49	51.35	90.77	91.31	83.31	92.54	91.06	86.31
CTMFNet (Song et al. 2022)	ResNet50	53.83	63.35	54.77	64.36	30.58	39.17	49.78	50.84	91.09	91.63	83.85	92.98	91.49	87.09
SAM-RS (Ma et al. 2024)	ResNet18	47.92	61.25	54.34	68.03	25.21	40.13	48.21	49.30	90.09	91.24	82.88	92.22	90.40	85.88
LOGCAN++ (Ma et al. 2025)	ResNet50	53.47	66.04	56.91	71.58	24.61	43.33	49.93	52.27	91.35	91.78	84.28	92.93	91.45	87.00
GSNet (Ye, Zhuge, and Zhang 2025)	CLIP-ViT-B	52.57	60.35	50.77	68.20	21.33	44.01	49.87	49.59	90.74	91.29	83.44	92.44	90.78	87.30
UNetFormer (Wang et al. 2022b)	ResNet18	52.89	61.73	53.44	65.76	33.18	43.29	46.57	50.98	89.96	90.86	81.98	92.17	90.68	85.71
	ResNet34	54.03	63.09	54.48	68.24	30.02	40.50	49.37	51.39	90.29	91.14	82.51	92.38	91.01	86.06
	ResNet50	52.55	61.36	52.11	69.62	30.22	41.94	52.87	51.52	90.59	91.29	83.11	92.57	91.03	86.39
UNetFormer+GAAR	ResNet18	53.73	63.67	53.93	67.61	37.10	43.91	49.65	52.80	90.25	91.12	82.46	92.59	91.09	86.41
	△	+0.84	+1.94	+0.49	+1.85	+3.92	+0.62	+3.08	+1.82	+0.29	+0.26	+0.48	+0.42	+0.41	+0.70
	ResNet34	55.18	64.69	55.38	69.70	35.42	42.58	52.32	53.61	90.52	91.36	82.90	92.65	91.19	86.52
	△	+1.15	+1.60	+0.90	+1.46	+5.40	+2.08	+2.95	+2.22	+0.23	+0.22	+0.39	+0.27	+0.18	+0.46
	ResNet50	53.95	63.98	52.96	70.35	31.90	45.67	59.53	54.05	91.15	91.71	83.96	92.83	91.37	86.83
	△	+1.40	+2.62	+0.85	+0.73	+1.68	+3.73	+6.66	+2.53	+0.56	+0.42	+0.85	+0.26	+0.34	+0.44
DC-Swin (Wang et al. 2022a)	Swin-T	53.05	62.26	56.02	69.28	36.09	42.67	51.06	52.92	90.01	91.21	82.09	91.98	90.12	85.31
	Swin-S	54.16	65.59	56.23	73.04	33.73	42.61	54.69	54.29	90.64	91.63	83.10	92.67	91.20	87.11
	Swin-B	54.57	61.73	57.31	69.33	30.03	42.80	52.10	52.55	90.53	91.74	83.40	92.85	91.34	86.87
DC-Swin+GAAR	Swin-T	55.09	66.43	56.78	69.64	37.03	42.95	51.22	54.16	90.23	91.33	82.45	92.33	90.86	85.95
	△	+2.04	+4.17	+0.76	+0.36	+0.94	+0.28	+0.16	+1.24	+0.22	+0.12	+0.36	+0.35	+0.74	+0.64
	Swin-S	54.48	65.84	57.27	73.80	36.62	45.86	54.57	**55.49**	91.02	91.73	83.73	93.07	91.67	87.92
	△	+0.32	+0.25	+1.04	+0.76	+2.89	+3.25	-0.12	+1.20	+0.38	+0.10	+0.63	+0.40	+0.47	+0.81
	Swin-B	54.85	66.57	57.41	68.74	31.30	45.17	52.28	53.76	90.94	91.80	83.60	93.01	91.63	87.13
	△	+0.28	+4.84	+0.10	-0.59	+1.27	+2.37	+0.18	+1.21	+0.41	+0.06	+0.20	+0.16	+0.29	+0.26
LSKNet (Li et al. 2024)	LSKNet-T	54.83	60.60	55.32	67.75	27.72	42.46	56.34	52.15	90.94	91.52	83.61	92.25	91.12	86.07
	LSKNet-S	54.05	61.62	55.03	70.07	30.52	39.80	54.24	52.19	91.01	91.70	83.72	93.06	91.98	87.19
LSKNet+GAAR	LSKNet-T	56.33	63.58	57.59	67.42	32.91	44.11	56.61	54.08	91.38	91.95	84.54	92.67	91.32	86.55
	△	+1.50	+2.98	+2.27	-0.33	+5.19	+1.65	+0.26	+1.93	+0.44	+0.43	+0.93	+0.42	+0.20	+0.48
	LSKNet-S	54.57	67.08	56.08	72.44	29.62	41.54	58.58	54.27	**91.70**	**92.07**	**84.59**	**93.54**	**92.59**	**88.15**
	△	+0.52	+5.46	+1.05	+2.37	-0.90	+1.71	+4.34	+2.08	+0.69	+0.37	+0.87	+0.48	+0.61	+0.96

Table 1: Comparison with SOTA methods on the LoveDA, Vaihingen, and Potsdam datasets (partial). △ indicates the performance change from the base model to its GAAR-enhanced variant. Best results are highlighted in bold.

Method	ResNet18	ResNet34	ResNet50
UNetFormer	50.98	51.39	51.52
w/o FCW Module	43.73	30.66	42.25
w/o Hyperbolic Attention	51.32	52.21	52.20
w/o Euclidean Attention	50.71	52.55	52.71
w/o GADR	50.66	52.66	52.00
w/o Adaptive Router	50.48	52.20	53.41
w/o Temperature Annealing	51.33	53.04	53.35
UNetFormer + GAAR	**52.80**	**53.61**	**54.05**

Table 2: Ablation results of GAAR on UNetFormer with ResNet18/34/50 backbones on the LoveDA dataset.

moderate degradation, while removing all of them (matching the base UNetFormer) produces a consistent performance gap. GADR enforces geometric consistency, stochastic routing provides regularization through sampling noise, and annealing stabilizes training. Together, these mechanisms are key to GAAR's robustness and generalization.

Hyperparameter Sensitivity We analyze GAAR's sensitivity to three key hyperparameters: the annealing temperature τ, the regularization weight λ, and the number of attention heads per branch. As shown in Tables 3–5, the best performance is obtained with $\tau \in [0.1, 1.0]$, $\lambda = 0.005$, and an asymmetric head allocation of 4 Hyperbolic heads and 8 Euclidean heads. This asymmetry likely reflects the different capacities of the two spaces: Hyperbolic geometry supports hierarchical abstraction, whereas Euclidean geometry captures fine-grained spatial detail.

Temperature (Init, Min)	ResNet18	ResNet34	ResNet50
Init=0.5, Min=0.1	51.76	52.46	53.69
Init=1.0, Min=0.1	**52.80**	**53.61**	**54.05**
Init=2.0, Min=0.1	51.80	50.34	53.24
Init=0.5, Min=0.01	51.57	52.79	52.37
Init=1.0, Min=0.01	51.61	52.75	51.89
Init=2.0, Min=0.01	51.09	52.22	52.01

Table 3: Effect of temperature annealing on GAAR performance using ResNet18/34/50 backbones.

λ	ResNet18	ResNet34	ResNet50
0.1	51.47	52.49	51.19
0.01	52.01	52.77	53.15
0.001	52.65	53.54	53.98
0.005	**52.80**	**53.61**	**54.05**

Table 4: Effect of the GADR hyperparameter λ on GAAR performance with ResNet18/34/50 backbones.

Qualitative Representation Analysis To visualize how GAAR improves feature embedding quality, we apply t-SNE (Van der Maaten and Hinton 2008) to training and test set representations (Figures 4 and 5). GAAR produces more compact intra-class clusters and larger inter-class separation compared to the UNetFormer baseline (Figure 4), indicating better semantic disentanglement. Additionally, test-time features exhibit stronger structural coherence (Figure 5), suggesting improved generalization. We attribute this to the

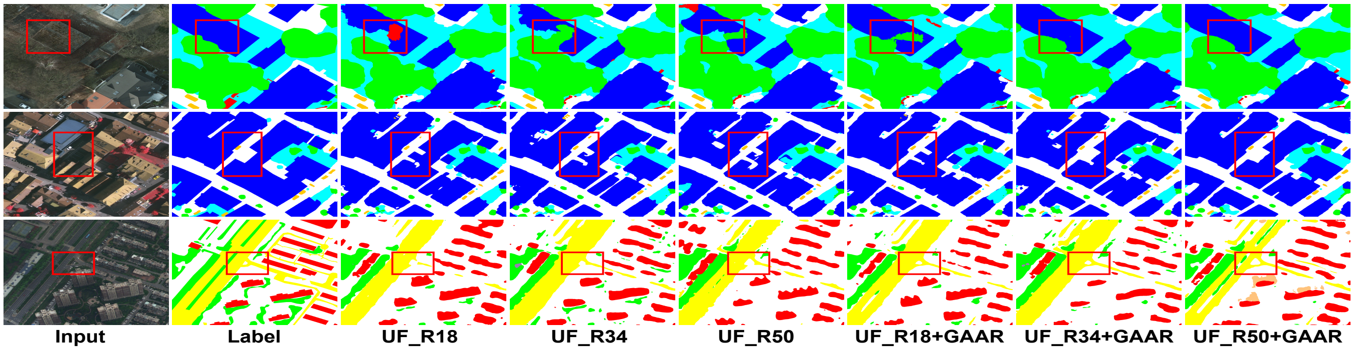

Figure 3: Visual comparison of UNetFormer with *vs.* without GAAR on Potsdam, Vaihingen, and LoveDA. Highlighted regions show enhanced semantic delineation.

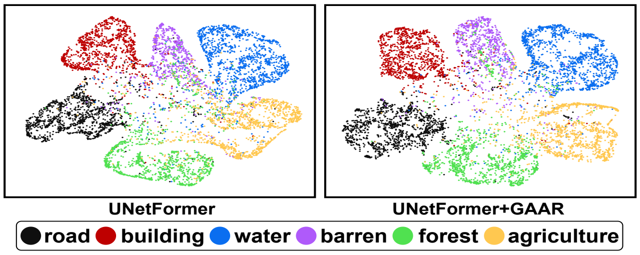

Figure 4: t-SNE visualization of segmentation features on LoveDA. GAAR yields tighter intra-class clusters and clearer inter-class separation than the baseline.

Head Configuration	ResNet18	ResNet34	ResNet50
HYP=4, EUC=4	51.04	52.84	53.25
HYP=4, EUC=8	**52.80**	**53.61**	**54.05**
HYP=8, EUC=4	51.53	52.94	53.47
HYP=8, EUC=8	51.33	53.20	53.60

Table 5: Effect of attention head count in GAAR using ResNet18/34/50 backbones. HYP and EUC denote the number of heads in Hyperbolic and Euclidean branches.

Gumbel sampling mechanism, which acts as an implicit regularizer that encourages robustness to input perturbations. While this can slightly reduce training accuracy, the improved test-time performance validates its effectiveness.

Efficiency–Accuracy Trade-Off Table 6 highlights GAAR's key contribution: a sparsely mixed Euclidean-Hyperbolic routing scheme that secures the best segmentation accuracy (mIoU=53.17) while staying computationally practical. Relative to the fully Hyperbolic baseline, our hybrid design is over four times faster (30.31 vs 7.45 FPS) and requires about 7.6 GB less memory (10.29 vs 17.86 GB). Although Euclidean-only U-Net offers higher raw throughput, the proposed hybrid still operates above 30 FPS—sufficient for real-time inference—yet raises mIoU by nearly five points. Ablations with GAAR_EUC and GAAR_HYP confirm that routing alone cannot match this accuracy; only the joint use of both geometries delivers consistent gains.

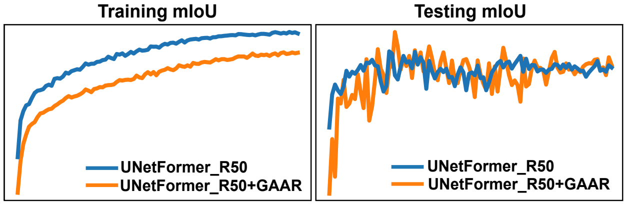

Figure 5: Impact of GAAR on training and testing performance. GAAR reduces overfitting through geometry-aware regularization and improves generalization.

Model Variant	FPS ↑	Memory (GB) ↓	mIoU ↑
(1) Baselines: Single-Space Representations			
Base_EUC (Euclidean Only)	**89.46**	**6.22**	48.23
Base_HYP (Hyperbolic Only)	7.45	17.86	**49.07**
(2) With Geometry-Aware Routing (GAAR Variants)			
GAAR_EUC (Pure Euclidean GAAR)	**52.14**	**8.37**	50.14
GAAR_HYP (Pure Hyperbolic GAAR)	13.76	12.06	51.22
GAAR (Hybrid Routing)	30.31	10.29	**53.17**

Table 6: Performance–efficiency of GAAR vs. Euclidean- and Hyperbolic-only U-Net variants (Ronneberger, Fischer, and Brox 2015). Hybrid routing attains the top mIoU, runs faster, and uses less memory than the fully Hyperbolic model, confirming the value of our hybrid routing approach.

Conclusions

We present GAAR, a geometry-aware routing module that adaptively assigns high-level features to Euclidean or Hyperbolic spaces based on structural complexity. Motivated by fractal dimension disparities among geospatial objects, GAAR addresses the representational limitations of Euclidean models while avoiding the computational burden of fully Hyperbolic architectures. Integrated with the GADR, our method enables sparse, structure-aligned and near-deterministic routing, supporting efficient end-to-end training on modern hardware. Extensive experiments across three RSISS benchmarks and diverse backbones demonstrate that GAAR consistently enhances segmentation accuracy, boundary precision, and topological coherence—offering a scalable and generalizable solution to long-standing geometric modeling challenges in remote sensing.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Fujian Province under Grants 2020J01571 and 2016J01281; in part by the Science and Technology Innovation Special Fund of Fujian Agriculture and Forestry University under Grant KFB25036A; and in part by the Research Fund of Fujian University of Technology under Grant GY-Z23210.

References

Chen, B.; Peng, W.; Cao, X.; and Röning, J. 2024. Hyperbolic Uncertainty Aware Semantic Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 25: 1275–1290.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2018a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834–848.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018b. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 833–851.

Dang, L.; Nie, Y.; Long, C.; Zhang, Q.; and Li, G. 2022. Diverse Human Motion Prediction via Gumbel-Softmax Sampling from an Auxiliary Space. In *Proceedings of the 30th ACM International Conference on Multimedia*, 5162–5171.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 1–21.

Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; and Lu, H. 2019. Dual Attention Network for Scene Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3146–3154.

Ganea, O.-E.; Bécigneul, G.; and Hofmann, T. 2018. Hyperbolic Neural Networks. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, 1–11.

GhadimiAtigh, M.; Schoep, J.; Acar, E.; van Noord, N.; and Mettes, P. 2022. Hyperbolic Image Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4453–4462.

Griffiths, D.; and Boehm, J. 2019. Improving Public Data for Building Segmentation from Convolutional Neural Networks (CNNs) for Fused Airborne Lidar and Image Data Using Active Contours. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154: 70–83.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Khrulkov, V.; Mirvakhabova, L.; Ustinova, E.; Oseledets, I.; and Lempitsky, V. 2020. Hyperbolic Image Embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6418–6428.

Kirillov, A.; Wu, Y.; He, K.; and Girshick, R. 2020. PointRend: Image Segmentation as Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9799–9808.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep Learning. *Nature*, 521: 436–444.

Li, R.; Zheng, S.; Duan, C.; Wang, L.; and Zhang, C. 2022. Land Cover Classification from Remote Sensing Images Based on Multi-Scale Fully Convolutional Network. *Geospatial Information Science*, 25(2): 278–294.

Li, R.; Zheng, S.; Zhang, C.; Duan, C.; Su, J.; and Wang, L. 2021. Multiattention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5607713.

Li, Y.; Hou, Q.; Zheng, Z.; Cheng, M.-M.; Yang, J.; and Li, X. 2023. Large Selective Kernel Network for Remote Sensing Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 16794–16805.

Li, Y.; Li, X.; Dai, Y.; Hou, Q.; Liu, L.; Liu, Y.; Cheng, M.-M.; and Yang, J. 2024. LSKNet: A Foundation Lightweight Backbone for Remote Sensing. *International Journal of Computer Vision*, 133: 1410–1431.

Liu, Y.; He, Z.; and Han, K. 2025. Hyperbolic Category Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9891–9900.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10012–10022.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440.

Luo, Z.; Yang, W.; Yuan, Y.; Gou, R.; and Li, X. 2024. Semantic Segmentation of Agricultural Images: A Survey. *Information Processing in Agriculture*, 11: 172–186.

Ma, X.; Lian, R.; Wu, Z.; Guo, H.; Yang, F.; Ma, M.; Wu, S.; Du, Z.; Zhang, W.; and Song, S. 2025. LOGCAN++: Adaptive Local-Global Class-Aware Network for Semantic Segmentation of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 4404216.

Ma, X.; Wu, Q.; Zhao, X.; Zhang, X.; Pun, M.-O.; and Huang, B. 2024. SAM-Assisted Remote Sensing Imagery Semantic Segmentation with Object and Boundary Constraints. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 5636916.

Peng, W.; Varanka, T.; Mostafa, A.; Shi, H.; and Zhao, G. 2022. Hyperbolic Deep Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 10023–10044.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38 th International Conference on Machine Learning (PMLR)*, 8748–8763.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241.

Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Benitez, S.; and Breitkopf, U. 2020. International Society for Photogrammetry and Remote Sensing, 2D Semantic Labeling Contest. *Accessed: Mar. 1, 2025*.

Samie, A.; Abbas, A.; Azeem, M. M.; Hamid, S.; Iqbal, M. A.; Hasan, S. S.; and Deng, X. 2020. Examining the Impacts of Future Land Use/Land Cover Changes on Climate in Punjab Province, Pakistan: Implications for Environmental Sustainability and Economic Growth. *Environmental Science and Pollution Research*, 27: 25415–25433.

Shen, J.; Hu, Q.; Feng, T.; Wang, X.; Cui, H.; Wu, S.; and Zhang, W. 2025. In2NeCT: Inter-Class and Intra-Class Neural Collapse Tuning for Semantic Segmentation of Imbalanced Remote Sensing Images. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 6814–6822.

Song, P.; Li, J.; An, Z.; Fan, H.; and Fan, L. 2022. CTMFNet: CNN and Transformer Multiscale Fusion Network of Remote Sensing Urban Scene Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5900314.

Strudel, R.; Garcia, R.; Laptev, I.; and Schmid, C. 2021. Segmenter: Transformer for Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7262–7272.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, 9: 2579–2605.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 6000–6010.

Wang, J.; Zheng, Z.; Ma, A.; Lu, X.; and Zhong, Y. 2021. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation. In *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS)*, 1–12.

Wang, L.; Li, R.; Duan, C.; Zhang, C.; Meng, X.; and Fang, S. 2022a. A Novel Transformer Based Semantic Segmentation Scheme for Fine-Resolution Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 19: 6506105.

Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; and Atkinson, P.-M. 2022b. UNetFormer: A UNet-Like Transformer for Efficient Semantic Segmentation of Remote Sensing Urban Scene Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190: 196–214.

Yan, J.; Luo, L.; Deng, C.; and Huang, H. 2021. Unsupervised Hyperbolic Metric Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12465–12474.

Yang, F.; Li, X.; and Shen, J. 2022. Nested Architecture Search for Point Cloud Semantic Segmentation. *IEEE Transactions on Image Processing*, 32: 2889–2900.

Ye, C.; Zhuge, Y.; and Zhang, P. 2025. Towards Open-Vocabulary Remote Sensing Image Semantic Segmentation. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 9436–9444.

Zhang, Z.; Ran, R.; Tian, C.; Zhou, H.; Li, X.; Yang, F.; and Jiao, Z. 2023. Self-aware and Cross-sample Prototypical Learning for Semi-supervised Medical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 192–201.

Zhao, J.; Li, X.; Yang, F.; Zhai, Q.; Luo, A.; Zhao, Y.; Cheng, H.; and Fu, H. 2025a. MExD: An Expert-Infused Diffusion Model for Whole-Slide Image Classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20789–20799.

Zhao, J.; Yang, F.; Li, X.; Jiao, Z.; Zhai, Q.; Li, X.; Wu, D.; Fu, H.; and Cheng, H. 2025b. SegMIC: A Universal Model for Medical Image Segmentation Through In-Context Learning. *Pattern Recognition*, 171: 112179.