Do We Need Perfect Data? Leveraging Noise for Domain Generalized Segmentation

Taeyeong Kim, SeungJoon Lee, Jung Uk Kim*, MyeongAh Cho*

Kyung Hee University, Republic of Korea
{rlaxo0511, diplomat3334, ju.kim, maycho}@khu.ac.kr

Abstract

Domain generalization in semantic segmentation faces challenges from domain shifts, particularly under adverse conditions. While diffusion-based data generation methods show promise, they introduce inherent misalignment between generated images and semantic masks. This paper presents **FLEX-Seg** (FLexible Edge eXploitation for Segmentation), a framework that transforms this limitation into an opportunity for robust learning. FLEX-Seg comprises three key components: (1) **Granular Adaptive Prototypes** that captures boundary characteristics across multiple scales, (2) **Uncertainty Boundary Emphasis** that dynamically adjusts learning emphasis based on prediction entropy, and (3) **Hardness-Aware Sampling** that progressively focuses on challenging examples. By leveraging inherent misalignment rather than enforcing strict alignment, FLEX-Seg learns robust representations while capturing rich stylistic variations. Experiments across five real-world datasets demonstrate consistent improvements over state-of-the-art methods, achieving 2.44% and 2.63% mIoU gains on ACDC and Dark Zurich. Our findings validate that adaptive strategies for handling imperfect synthetic data lead to superior domain generalization.

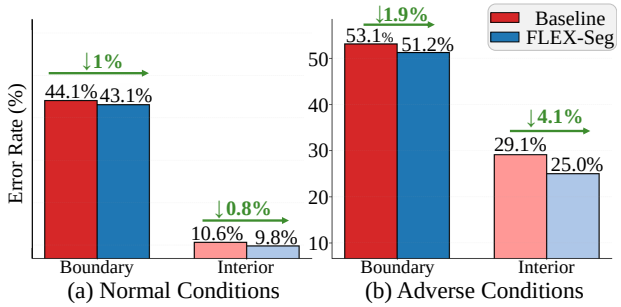


Figure 1: Error rate analysis of boundary vs. interior regions when training with diffusion-generated synthetic data. Under both (a) normal and (b) adverse conditions, boundary regions consistently exhibit higher error rates than interior regions, with the disparity becoming more pronounced in adverse conditions. **FLEX-Seg** effectively reduces errors in both regions across all scenarios, demonstrating that our boundary-focused approach is crucial for robust domain generalization.

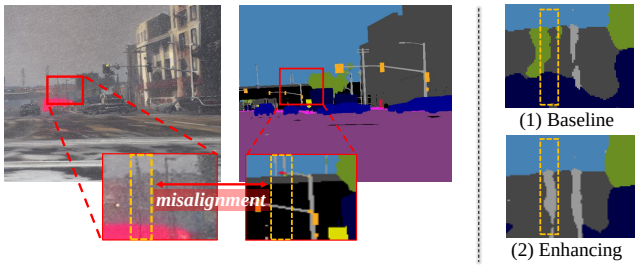
Introduction

Semantic segmentation, which assigns semantic labels to each pixel in an image, is fundamental to many computer vision applications, especially autonomous driving (Shelhamer, Long, and Darrell 2014; Chen et al. 2017; Badrinarayanan, Kendall, and Cipolla 2017). However, models trained on one domain often fail on unseen domains due to *domain shift* caused by variations in weather, lighting, and imaging conditions (Li et al. 2023a; Hoffman et al. 2018), creating a *domain gap* (Luo et al. 2019) that severely limits real-world deployment. While Domain Adaptive Semantic Segmentation addresses this by using unlabeled target domain data during training (Ganin and Lempitsky 2015; Hoffman et al. 2018), collecting such data is often impractical (Li et al. 2017a). This motivates Domain Generalized Semantic Segmentation (DGSS), which trains models to generalize to unseen domains using only source domain data (Zhou et al. 2022; Carlucci et al. 2019), eliminating any need for target domain data access (neither images nor labels). Thus, DGSS is more suitable for real-world applications where deployment environments are unknown or constantly changing.

Recent studies in DGSS have explored diverse approaches for learning domain-invariant features. Early methods focused on normalization techniques (Pan et al. 2018; Choi et al. 2021) and style randomization (Yue et al. 2019; Kim, Kim, and Kim 2023) to reduce domain-specific biases. Subsequently, data augmentation via image-to-image translation has been widely adopted to simulate diverse domain conditions (Li et al. 2023b; Huang et al. 2021). Recently, diffusion-based generative methods have set new benchmarks. Particularly, DGInStyle (Jia et al. 2024) leverages pre-trained latent diffusion models to synthesize diverse yet semantically consistent images, significantly improving generalization by exploiting large-scale generative priors, although they often introduce spatial misalignments at fine-grained levels such as object boundaries.

However, a fundamental challenge remains: accurately predicting object boundaries is crucial for semantic segmentation, as boundaries define the shape and structure of objects (Marmanis et al. 2018; Li et al. 2020). This challenge

*Corresponding author.



(a) Generated Image (b) Ground Truth (c) Prediction

Figure 2: Column 1: Generated image and corresponding ground truth mask. Column 2: Model predictions on (a) generated image using different approaches. **Better view in zoom.**

is particularly critical in DGSS, where models must generalize boundary representations to diverse boundary appearances across unseen domains without access to target domain data, making them more vulnerable to cross-domain variations in boundary visibility and structure. The inherent characteristics of adverse conditions—such as fog, snow, and nighttime scenarios—make this challenge even more severe, as these conditions naturally produce blurred boundaries and reduced visibility that obscure the clear delineation between objects (Lei et al. 2020; Narayanan, Rajendran, and Kimia 2021). As shown in Fig. 1, boundary regions exhibit significantly higher error rates than interior regions, with errors becoming even larger under adverse conditions. This issue is exacerbated when utilizing generated synthetic datasets for DGSS, where the generation process often fails to achieve *perfect pixel-wise alignment*. In contrast to real datasets—where semantic mask labels are typically derived from real images—synthetic data pipelines generate images from semantic masks. This reversed generation process inherently introduces spatial misalignments between the synthesized images and their associated labels. As illustrated in the first column of Fig. 2, fine-grained structures in the generated images often fail to align precisely with their semantic masks. Such inconsistencies can mislead the model into focusing on resolving local ambiguities rather than learning meaningful cross-domain variations.

Existing boundary-aware methods (Zhang et al. 2024; Borse et al. 2021; Ngoc et al. 2021; Liu et al. 2021) are fundamentally limited in handling the inherent misalignment present in synthetic data. These approaches are predicated on the assumption of perfect spatial correspondence between image structures and semantic mask annotations—an assumption that does not hold in synthetically generated datasets. Therefore, addressing the dual challenges of boundary precision from synthetic data misalignment requires a new approach that learns robust representations from imperfect alignments while capturing rich stylistic variations in less constrained regions.

To address these challenges, we propose **FLEX-Seg** (**FL**exible **E**dge e**X**ploitation for **Seg**mentation), a novel framework that transforms the inherent misalignment in synthetic data into an opportunity for learning more robust and domain-invariant representations. FLEX-Seg comprises three carefully designed components that synergistically balance precise boundary delineation with rich stylistic learning. (1) **Granular Adaptive Prototypes (GAP)** captures boundary characteristics at different thickness levels by organizing prototypes according to both semantic class and boundary granularity. By constructing a structured prototype bank across three granularity levels—from thin boundaries in distant objects to thick regions in nearby ones—GAP learns domain-invariant boundary representations while maintaining the natural scale variations of object boundaries. (2) **Uncertainty Boundary Emphasis (UBE)** modulates supervision strength based on prediction entropy. This mechanism dynamically amplifies the contribution of uncertain pixels—typically at misaligned boundaries and ambiguous regions—while maintaining standard gradients for confident predictions. By directing learning capacity toward these challenging areas, UBE enhances boundary discrimination without overfitting to synthetic data artifacts, improving robustness across diverse domains. (3) **Hardness-Aware Sampling (HAS)** optimizes training efficiency by progressively focusing on challenging examples. Through sigmoid decay scheduling that gradually transitions from random to loss-based sampling, HAS ensures efficient resource allocation to the most informative samples, particularly those with complex structures or adverse conditions.

Comprehensive domain generalization experiments across five real-world datasets demonstrate that FLEX-Seg consistently outperforms existing state-of-the-art methods. Our approach achieves significant improvements on challenging domains with adverse conditions, gaining 2.44% and 2.63% mIoU on ACDC (fog, rain, snow, and nighttime scenarios) and Dark Zurich (nighttime driving conditions) respectively. These results validate our hypothesis that leveraging inherent misalignment in synthetic data rather than enforcing strict alignment leads to learn more robust and transferable representations, ultimately leading to improved generalization across unseen domains.

Related Work

Dataset Generation for Domain Generalization

Recent approaches improve training data quality using diffusion models for diverse synthetic generation. CLOUDS (Benigmim et al. 2024) combines foundation models for enhanced diversity in domain generalization. ALDM (Li et al. 2024) introduces adversarial supervision to preserve layout fidelity in layout-to-image diffusion models. DGInStyle (Jia et al. 2024) uses latent diffusion with style swapping to create semantically consistent images across domains. Despite these efforts to improve generation quality, these methods often cause boundary misalignment between generated images and semantic masks, an issue that remains underexplored; our approach leverages this misalignment for robust representations.

Domain Generalized Semantic Segmentation

Domain generalization in semantic segmentation aims to train models that perform well on unseen domains with-

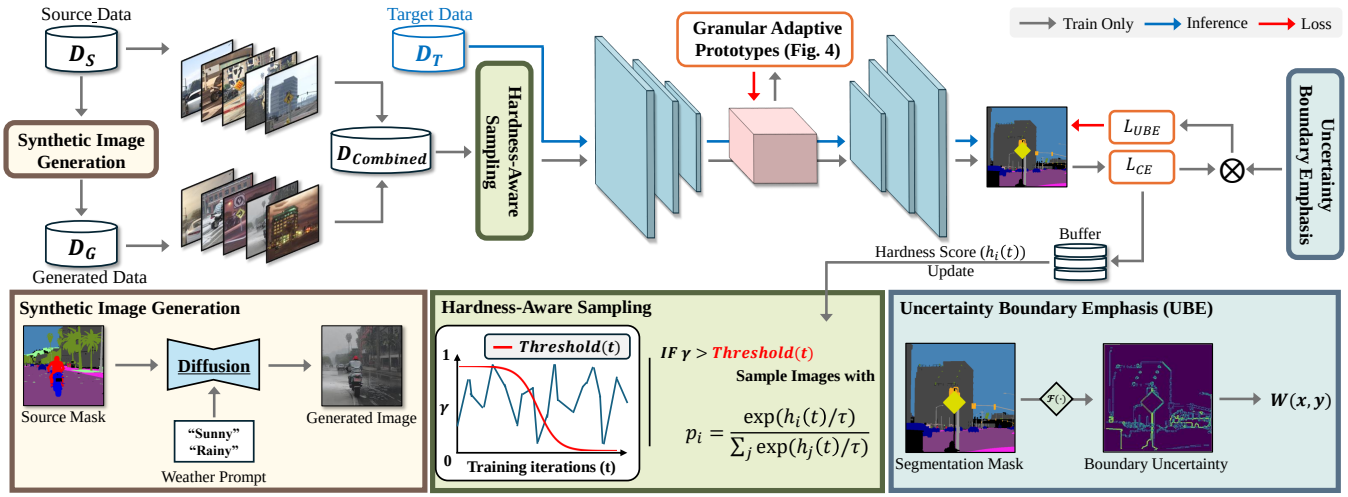


Figure 3: Overview of our FLEX-Seg framework. The framework integrates three key components: Granular Adaptive Prototypes for learning domain-invariant boundary representations, Uncertainty Boundary Emphasis for dynamically emphasizing challenging regions, and Hardness-Aware Sampling for efficient training on difficult examples. These components work synergistically to improve boundary precision across diverse domains.

out access to target data. Early methods used normalization techniques like instance normalization (Pan et al. 2018) and whitening transformation (Li et al. 2017b; Choi et al. 2021) to reduce domain-specific style variations. Domain randomization simulates diverse conditions via style transfer and augmentation (Yue et al. 2019; Kim, Kim, and Kim 2023). Recent advances leverage modern architectures like DAFormer (Hoyer, Dai, and Van Gool 2022) for Transformer-based training, HRDA (Hoyer, Dai, and Van Gool 2023) for multi-resolution fusion. Further extensions include language-guided approaches (Fahes et al. 2024) and semantic consistency learning (Niu et al. 2025).

Motivation

We first investigate the effect of different boundary handling strategies to address the misalignment issue between the synthesized images and their associated labels in synthetic training data. Since misalignment primarily manifests at object boundaries where generated textures fail to precisely follow semantic masks, we hypothesize that focusing on these critical regions could improve model robustness. The most intuitive approach is to apply stronger weights to boundary regions during training, forcing the model to pay more attention to these challenging areas despite their imperfect alignment. We identify boundary regions B through morphological operations on semantic masks:

$$B = \text{Dilate}(M, k_d) - \text{Erode}(M, k_e), \tag{1}$$

where M is the semantic mask, and k_d, k_e are dilation and erosion kernel sizes respectively.

We then incorporate enhanced weighting into the training loss to explicitly emphasize learning on boundary regions:

$$\mathcal{L}_{\text{boundary}} = \sum_{(x,y)} W(x,y) \cdot \mathcal{L}_{\text{CE}}(x,y), \tag{2}$$

where $W(x,y) = \alpha > 1$ if $(x,y) \in B$, and 1 otherwise. The parameter α adjusts the loss weight for boundary pixels, guiding the model to focus more on accurately capturing object boundaries.

As shown in the second column of Fig. 2, this simple boundary-enhancing approach effectively captures object characteristics even in misaligned synthetic data. The model trained with boundary emphasis ($\alpha = 5$) shows more robust predictions compared to the baseline, successfully delineating object structures despite the substantial misalignment between generated images and semantic masks.

This observation inspired us to design more sophisticated modules that not only emphasize boundaries but also adapt to their inherent uncertainty and multi-scale characteristics. Rather than using a fixed weight α, our GAP module learns boundary representations across multiple granularities, while UBE dynamically adjusts weights based on prediction uncertainty. Additional motivation experiments exploring various boundary weighting strategies (ignore, reduce, threshold) can be found in the supplementary material.

Method

DG for Semantic Segmentation

The goal of domain generalization in semantic segmentation is to learn a model using only labeled source domain data that can perform well on unseen target domains. Formally, let the source domain be denoted by $\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{N_S}$, where x_i^S represents an image in the source domain and y_i^S is its pixel-level semantic label. Our objective is to learn a segmentation model that generalizes to unseen target domains $\mathcal{D}_T = \{x_i^T\}_{i=1}^{N_T}$, where x_i^T denotes an image in the target domain and labels are unavailable during training, reflecting the typical domain generalization setup.

FLEX-Seg Framework

The FLEX-Seg framework addresses boundary precision challenges in diffusion-based domain generalization through two sophisticated modules: Granular Adaptive Prototypes (GAP) and Uncertainty Boundary Emphasis (UBE), alongside Hardness-Aware Sampling (HAS). Fig. 3 illustrates the comprehensive architecture of our approach.

Initially, diverse synthetic images are generated using diffusion models on source dataset $\mathcal{D}_S$ and its semantic masks, forming augmented dataset $\mathcal{D}_G$ and combining it with $\mathcal{D}_S$ to create unified training corpus $\mathcal{D}_{combined} = \mathcal{D}_S \cup \mathcal{D}_G$.

The training process incorporates a dual-pathway boundary refinement strategy. GAP constructs a two-dimensional prototype bank that captures boundary characteristics across class semantics and granular intensity levels, enabling domain-invariant representation learning through contrastive alignment, while UBE dynamically adjusts pixel-wise loss contributions based on prediction uncertainty, providing adaptive emphasis on challenging boundary regions without manual hyperparameter tuning.

Integration of prototype-based learning and uncertainty-driven weighting provides a robust foundation for domain generalization, effectively handling diverse visual conditions and complex boundary structures across domains.

Granular Adaptive Prototypes (GAP)

The GAP module addresses domain-invariant boundary representation learning through a novel two-dimensional coordinate system, as illustrated in Fig. 4. This approach stems from the inherent scale variability of semantic boundaries, where distant small objects exhibit thin boundaries while nearby large objects present thick boundary regions. Additionally, boundary pixels exhibit two distinct types of variations: geometric variations (how thick or thin the boundary appears) and stylistic variations (how the boundary appears under different environmental conditions).

Class-Shape Token Coordinate System. We formalize this concept by decomposing boundary characteristics into orthogonal dimensions. Each boundary pixel p_i is represented as a point in a coordinate space (c_i, g_i), where the class token c_i captures the semantic class identity, and the shape token g_i encodes geometric attributes such as boundary thickness.

To extract multi-granular boundary representations, we downsample the ground truth mask M from resolution (H, W) to match feature maps at (H_f, W_f) using stride $s = H/H_f$ to obtain M_d. We then generate three boundary masks through morphological operations:

$$B_g = \text{Dilate}(M_d, k_g) \ominus \text{Erode}(M_d, k_g) \tag{3}$$

where k_g represents kernel sizes for generating thin, medium, and thick boundary granularities respectively.

This decomposition enables systematic analysis of boundary characteristics by separating semantic and geometric properties that should remain consistent across domains from stylistic variations due to environmental conditions.

Prototype Bank Construction. We construct a prototype bank $\mathcal{P} = \{p_{c,g}\}$ where $p_{c,g} \in \mathbb{R}^{256}$ represents the prototype for class c with granularity g, resulting in $C \times 3$ prototypes (C classes × 3 granularities). The prototype bank serves as a structured memory that captures representative boundary features for each class-granularity combination, enabling consistent boundary representations across different domains.

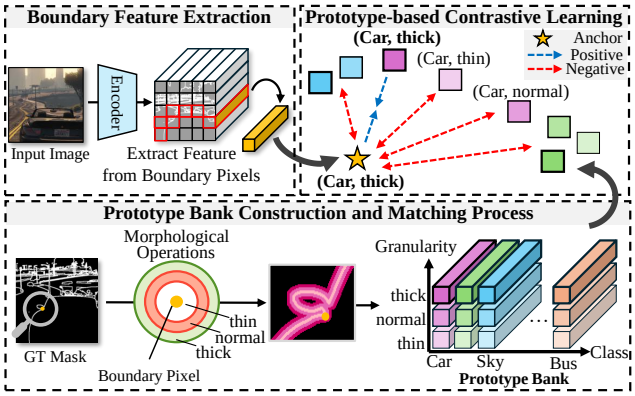


Figure 4: Illustration of GAP. Boundary features are extracted at multiple granularities, assigned to corresponding class-granularity prototypes, and refined through contrastive learning to achieve domain-invariant representations.

Prototypes are updated via pixel-wise momentum updates during training:

$$p_{c,g} \leftarrow m \cdot p_{c,g} + (1 - m) \cdot f_{c,g} \tag{4}$$

where $m \in [0, 1]$ is the momentum factor, and $f_{c,g}$ is the feature vector of a boundary pixel belonging to class c with granularity g. This pixel-wise update strategy allows for fine-grained prototype refinement while maintaining stability through the momentum mechanism.

Prototype-based Contrastive Learning. The InfoNCE loss enforces that boundary pixels with identical class-granularity characteristics converge regardless of domain origin. To address class imbalance, we incorporate adaptive weighting based on prototype exposure frequency:

$$\mathcal{L}_{GAP} = -\frac{1}{N} \sum_{i=1}^{N} w_{c_i,g_i} \cdot \log \frac{e^{\langle f_i, p_{c_i,g_i} \rangle / \tau}}{\sum_{(c',g') \in \mathcal{P}} e^{\langle f_i, p_{c',g'} \rangle / \tau}} \tag{5}$$

where τ is the temperature parameter, N is the number of boundary pixels in the batch, $\mathcal{P}$ represents all class-granularity combinations in the prototype bank, and $\langle \cdot, \cdot \rangle$ denotes the cosine similarity. The weight w_{c_i,g_i} balances the contribution of each class-granularity combination.

For each prototype $p_{c,g}$, we maintain its update frequency $u_{c,g}$. The imbalance-aware weight is computed as:

$$w_{c,g} = \frac{\max(u) + 1}{u_{c,g} + 1} \cdot \frac{1}{Z} \tag{6}$$

where $\max(u)$ denotes the highest update frequency among all prototypes, and Z is a normalization factor ensuring $\max(w_{c,g}) = 1$. This formulation assigns higher weights to under-represented class-granularity combinations, effectively balancing the learning across all boundary types.

This comprehensive approach enables domain-invariant boundary learning while addressing the natural imbalance in boundary occurrence frequencies.

Uncertainty Boundary Emphasis (UBE)

The UBE module dynamically emphasizes challenging boundary regions by leveraging prediction uncertainty, eliminating the need for manual hyperparameter tuning across diverse domains. Unlike fixed weighting schemes, this approach automatically adapts to varying difficulty levels where nighttime or adverse weather conditions present more uncertain boundary predictions.

Entropy-based Dynamic Weighting. We compute prediction entropy as an uncertainty indicator, where higher values correspond to ambiguous boundary regions. For each pixel location (x, y), the entropy is calculated as $H_{x,y} = -\sum_{c=1}^{C} p_c(x, y) \log p_c(x, y)$, where $p_c(x, y)$ is the predicted probability for class c at location (x, y). The adaptive weighting is applied exclusively to boundary regions while maintaining unit weight for interior pixels:

$$w(x,y) = \begin{cases} 1 + \alpha \cdot \text{sigmoid}\left(\frac{H_{x,y} - \mu_H}{\sigma_H + \epsilon}\right), & \text{if } (x,y) \in B \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

where B represents the boundary mask, μ_H and σ_H are batch-wise entropy statistics within boundary regions, and α controls maximum weight amplification.

The uncertainty-adaptive weights are incorporated into the cross-entropy loss:

$$\mathcal{L}_{UBE} = \frac{1}{N} \sum_{(x,y)} w(x,y) \cdot \mathcal{L}_{CE}(x,y) \tag{8}$$

where N is the total number of pixels and $\mathcal{L}_{CE}(x, y)$ is the cross-entropy loss at pixel location (x, y). This mechanism automatically emphasizes difficult boundary regions while preserving stable learning for confident predictions, requiring only a single stable hyperparameter α across different environmental conditions.

Loss Function Integration. The complete training objective combines the UBE loss and GAP loss for comprehensive boundary refinement:

$$\mathcal{L}_{total} = \mathcal{L}_{UBE} + \lambda_{gap} \cdot \mathcal{L}_{GAP} \tag{9}$$

While GAP enables domain-invariant boundary learning through prototype-based contrastive learning, UBE provides adaptive emphasis on uncertain regions. Together, these mechanisms create a synergistic effect: GAP ensures consistent boundary representations across domains by learning from multi-granular prototypes, while UBE dynamically adjusts the learning focus based on prediction confidence. This integrated approach allows the model to effectively handle both the inherent misalignment in synthetic data and the varying difficulty levels across different environmental conditions, resulting in more robust domain generalization performance.

Hardness-Aware Sampling (HAS)

To optimize training efficiency, we employ HAS that dynamically prioritizes difficult examples during training. Our approach balances exploration and exploitation by gradually transitioning from random to loss-based sampling as training progresses.

For each training image i, we maintain a hardness score $h_i(t)$ reflecting its difficulty level at updating step t. This score is updated periodically using exponential moving average (EMA):

$$h_i(t) = \beta \cdot h_i(t-1) + (1-\beta) \cdot L\big(f_\theta(x_i), y_i\big) \tag{10}$$

where L represents the accumulated loss value, f_θ is the segmentation model with parameters θ, and $\beta \in (0, 1)$ is the EMA decay factor. Periodic updates provide more stable hardness estimates compared to per-iteration updates.

The sampling strategy employs an adaptive threshold mechanism with sigmoid decay:

$$\text{threshold}(t) = \frac{1}{1 + e^{k(t-m)}} \tag{11}$$

where k controls the steepness of decay and m is the transition midpoint. At each training iteration, we generate a random value $r \in [0, 1]$. If $r > \text{threshold}(t)$, we perform loss-based sampling; otherwise, we sample randomly from the training set.

When loss-based sampling is selected, images are sampled with probability proportional to their hardness scores:

$$p_i = \frac{\exp\big(h_i(t)/\tau\big)}{\sum_j \exp\big(h_j(t)/\tau\big)} \tag{12}$$

where τ is a temperature parameter controlling the sampling distribution sharpness. Higher hardness scores result in higher sampling probabilities, directing attention toward challenging cases such as adverse weather conditions or complex boundary structures.

This design ensures training stability in early stages when loss estimates may be unreliable, while progressively focusing on the most informative examples as the model matures. The result is enhanced domain generalization without requiring additional training data or model complexity.

Experiments

Experimental Settings and Implementation Details

Datasets. Following standard practice in semantic segmentation domain generalization (Hoyer, Dai, and Van Gool 2023, 2022), we use **GTA** (Richter et al. 2016) as our source dataset, containing 24,966 synthetic images with 19-class annotations. For evaluation, we employ five real-world driving datasets spanning diverse conditions. We assess performance on **Cityscapes (CS)** (Cordts et al. 2016) (500 validation images from German cities), **BDD100K (BDD)** (Yu et al. 2020) (1,000 images from various US locations), and **Mapillary Vistas (MV)** (Neuhold et al. 2017) (2,000 images from worldwide locations). For challenging conditions, we include **ACDC** (Sakaridis, Dai, and Van Gool 2021) (406

Method	ACDC	DZ	Avg2	CS	BDD	MV	Avg5
				ResNet-101			
DRPC (Yue et al. 2019)	–	–	–	42.53	38.72	38.05	–
FSDR (Huang et al. 2021)	24.77	9.66	17.22	44.80	41.20	43.40	32.77
GTR (Peng et al. 2021)	–	–	–	43.70	39.60	39.10	–
SAN-SAW (Peng et al. 2022)	–	–	–	45.33	41.18	40.77	–
AdvStyle (Zhong et al. 2022)	–	–	–	44.51	39.27	43.48	–
SHADE (Zhao et al. 2022)	29.06	8.01	18.54	46.66	43.66	45.50	34.58
FAMix* (Fahes et al. 2024)	32.74	–	–	48.15	**45.61**	<u>52.11</u>	–
SCSD* (Niu et al. 2025)	<u>35.66</u>	–	–	**51.72**	<u>44.67</u>	**56.98**	–
HRDA (Hoyer et al. 2023)	26.08	7.80	16.94	39.63	38.69	42.21	30.88
+ DGInStyle (Jia et al. 2024)	34.19	16.16	25.18	46.89	42.81	50.19	38.05
++ FLEX-Seg (Ours)	**36.27**	**19.20**	**27.74**	<u>48.32</u>	44.03	51.13	**39.79**
				MiT-B5			
DAFormer (Hoyer et al. 2022)	38.25	17.45	27.85	52.65	47.89	54.66	42.18
+ DGInStyle (Jia et al. 2024)	44.04	25.58	34.81	55.31	50.82	56.62	46.47
++ FLEX-Seg (Ours)	**46.56**	**29.51**	**38.04**	**56.84**	**52.06**	**57.93**	**48.58**
HRDA (Hoyer et al. 2023)	44.04	20.97	32.51	57.41	49.11	61.16	46.54
+ DGInStyle (Jia et al. 2024)	46.07	25.53	35.80	58.63	52.25	**62.47**	48.99
++ FLEX-Seg (Ours)	**48.51**	**28.16**	**38.34**	**59.49**	**52.48**	61.71	**50.07**

Table 1: Domain generalization performance with GTA as source domain (mIoU ↑ in %). We compare our FLEX-Seg against state-of-the-art methods across both challenging conditions (ACDC, DZ) and standard datasets (CS, BDD, MV). For fair comparison, results are reported using both ResNet-101 and MiT-B5 backbones. The symbols + and ++ indicate incremental improvements, where ++ denotes methods built upon +. * means the method using ResNet-50 as the backbone, which is initialized with CLIP pretrained weights. We emphasize **best** and <u>second best</u> results.

images under fog, night, rain, and snow) and **Dark Zurich (DZ)** (Sakaridis, Dai, and Gool 2019) (50 nighttime images). This diverse collection of datasets enables thorough evaluation of our model's generalization capabilities across both standard and adverse driving conditions.

Implementation Details. For image generation, following (Jia et al. 2024), we implement Rare Class Sampling (RCS) with a probability threshold of $T = 0.01$ to address class imbalance issues in the generated dataset. Specifically, semantic masks from the GTA dataset are selected using RCS, cropped to 512×512 patches centered around rare-class regions, and used to generate 10,000 diverse images covering both standard and challenging weather conditions (fog, rain, snow, and nighttime). For training, we combine these generated images with an additional 6,000 images selected from the original GTA dataset based on rare-class criteria, further ensuring balanced class representation. Our FLEX-Seg framework maintains a prototype bank of size $C \times 3 \times 256$ for GAP, where prototypes are updated using momentum factor $m = 0.99$. Our HAS strategy maintains a hardness score for each training image, updated every 50 iterations using an exponential moving average (EMA) with decay factor $\beta = 0.9$. All experiments were conducted on single GPUs: RTX 3090 for DAFormer and A6000 Ada for HRDA. Other training settings (e.g., 3 random runs) follow the configurations in (Hoyer, Dai, and Van Gool 2023).

Hyperparameters. Our experiments determined the following optimal settings: GAP contrastive temperature $\tau = 0.07$, UBE entropy amplification factor $\alpha = 3.0$, HAS sigmoid decay parameters $k = 0.05$ with sampling temperature $\tau_{HAS} = 1.0$, and GAP loss weight $\lambda_{gap} = 0.5$. Detailed hyperparameter sensitivity analysis can be found in supplementary material.

Comparison with State-of-the-Art

Table 1 presents comprehensive comparisons with state-of-the-art domain generalization methods. Our FLEX-Seg demonstrates consistent improvements across all evaluation settings, achieving new state-of-the-art performance.

Challenging Domains. On domains with adverse conditions (ACDC, DZ), FLEX-Seg shows significant gains. For ACDC, we achieve 46.56% mIoU with DAFormer (+2.52% over DGInStyle) and 48.51% with HRDA (+2.44% improvement). Notably, our method surpasses recent strong baselines FAMix (32.74%) and SCSD (35.66%), despite these methods using CLIP-pretrained ResNet-50. On Dark Zurich nighttime scenarios, our method reaches 29.51% and 28.16% mIoU for DAFormer and HRDA respectively, representing substantial improvements of +3.93% and +2.63%. These gains demonstrate that our approach enables accurate object delineation even under poor visibility, where precise boundary understanding becomes critical for distinguishing objects from their surroundings.

Standard Domains. The improvements extend to standard conditions as well. On Cityscapes, BDD100K, and Mapillary Vistas, FLEX-Seg consistently outperforms previous methods across both ResNet-101 and MiT-B5 architectures. These results demonstrate that our boundary-focused approach not only excels in adverse conditions but also enhances segmentation quality under normal driving scenarios,

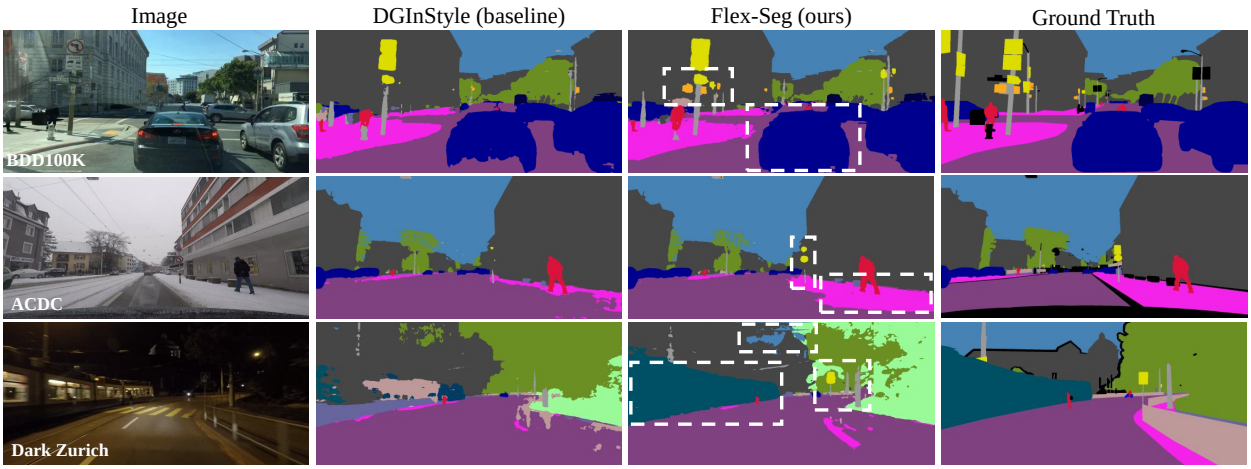


Figure 5: Qualitative comparison of segmentation results on target domains. From left to right: input images, predictions by HRDA trained with DGInStyle (Jia et al. 2024), predictions by HRDA trained with our FLEX-Seg, and ground truth.

GAP	UBE	HAS	Avg2	Avg3	Avg5
-	-	-	34.81	54.25	46.47
✓	-	-	36.33	55.26	47.69
-	✓	-	35.05	55.33	47.21
✓	✓	-	36.15	**56.07**	48.10
✓	✓	✓	**38.04**	55.61	**48.58**

Table 2: Ablation studies on key components across challenging (Avg2), standard (Avg3), and all domains (Avg5).

validating the general applicability of addressing boundary precision and uncertainty in synthetic training data.

Qualitative Analysis

Fig. 5 presents visual comparisons between baseline and our FLEX-Seg approach. The qualitative results reveal several key improvements: (1) More accurate boundary delineation, particularly visible in object contours such as vehicles and pedestrians; (2) Better handling of thin structures like poles and traffic signs, which are often misclassified by baseline methods; (3) Robust performance under adverse conditions including snow and nighttime scenarios, where visibility is severely limited. These visual improvements align with our quantitative results, confirming that FLEX-Seg's focus on multi-granular boundary learning and uncertainty-adaptive weighting effectively addresses the limitations of previous approaches. The enhanced precision is especially evident in challenging weather conditions, where our method successfully delineates objects despite snow occlusion or low-light environments, while baseline methods produce fragmented or inconsistent predictions.

Ablation Studies

Table 2 presents ablation studies validating the contribution of each component in FLEX-Seg. Starting from the DGInStyle baseline (34.81% Avg2), we observe that GAP alone provides substantial improvement (+1.52%), while UBE alone offers modest gains (+0.24%). Combining GAP and UBE yields further improvements, reaching 36.15% on challenging domains. The full framework with all three components achieves the best overall performance (38.04% Avg2, 48.58% Avg5), representing +3.23% improvement over baseline on challenging domains.

Interestingly, while GAP+UBE achieves the highest standard domain score (56.07% Avg3), adding HAS slightly reduces this (-0.46%) but significantly boosts challenging domain performance (+1.89%). This trade-off reflects HAS's adaptive sampling strategy, which prioritizes adverse examples with higher loss while maintaining balance to prevent standard domain degradation, ultimately enhancing overall robustness across diverse scenarios rather than single-environment optimization.

Conclusion

In this paper, we introduced **FLEX-Seg**, a comprehensive framework that transforms the inherent misalignment in synthetic data into an opportunity for robust domain generalization. Our approach comprises three synergistic components: Granular Adaptive Prototypes (GAP) for learning domain-invariant boundary representations across multiple scales, Uncertainty Boundary Emphasis (UBE) for dynamically emphasizing challenging regions based on prediction entropy, and Hardness-Aware Sampling (HAS) for efficient training on difficult examples. Extensive experiments across five diverse real-world datasets demonstrate that FLEX-Seg consistently outperforms existing state-of-the-art methods. Substantial gains on challenging domains (up to +3.93% on Dark Zurich) validate that leveraging boundary misalignment through adaptive strategies leads to superior generalization. Our work shows that precise boundary handling of imperfect synthetic data yields better generalization than striving for perfect alignment. Future directions include exploring adaptive prototype mechanisms and extending our framework to other dense prediction tasks where boundary precision is critical.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT)(RS-2024-00456589) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. RS-2025-02263277).

References

Badrinarayanan, V.; Kendall, A.; and Cipolla, R. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12): 2481–2495.

Benigmim, Y.; Roy, S.; Essid, S.; Kalogeiton, V.; and Lathuilière, S. 2024. Collaborating foundation models for domain generalized semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3108–3119.

Borse, S.; Wang, Y.; Zhang, Y.; and Porikli, F. 2021. Inverseform: A loss function for structured boundary-aware segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5901–5911.

Carlucci, F. M.; D'Innocente, A.; Bucci, S.; Caputo, B.; and Tommasi, T. 2019. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2229–2238.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. 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.

Choi, S.; Jung, S.; Yun, H.; Kim, J. T.; Kim, S.; and Choo, J. 2021. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11580–11590.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3213–3223.

Fahes, M.; Vu, T.-H.; Bursuc, A.; Pérez, P.; and De Charette, R. 2024. A simple recipe for language-guided domain generalized segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23428–23437.

Ganin, Y.; and Lempitsky, V. 2015. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, 1180–1189. PMLR.

Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A.; and Darrell, T. 2018. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, 1989–1998. Pmlr.

Hoyer, L.; Dai, D.; and Van Gool, L. 2022. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9924–9935.

Hoyer, L.; Dai, D.; and Van Gool, L. 2023. Domain adaptive and generalizable network architectures and training strategies for semantic image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1): 220–235.

Huang, J.; Guan, D.; Xiao, A.; and Lu, S. 2021. Fsdr: Frequency space domain randomization for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6891–6902.

Jia, Y.; Hoyer, L.; Huang, S.; Wang, T.; Van Gool, L.; Schindler, K.; and Obukhov, A. 2024. Dginstyle: Domain-generalizable semantic segmentation with image diffusion models and stylized semantic control. In *European Conference on Computer Vision*, 91–109. Springer.

Kim, S.; Kim, D.-h.; and Kim, H. 2023. Texture learning domain randomization for domain generalized segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 677–687.

Lei, Y.; Emaru, T.; Ravankar, A. A.; Kobayashi, Y.; and Wang, S. 2020. Semantic image segmentation on snow driving scenarios. In *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1094–1100. IEEE.

Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. M. 2017a. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, 5542–5550.

Li, J.; Xu, R.; Ma, J.; Zou, Q.; Ma, J.; and Yu, H. 2023a. Domain adaptive object detection for autonomous driving under foggy weather. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 612–622.

Li, X.; Li, X.; Zhang, L.; Cheng, G.; Shi, J.; Lin, Z.; Tan, S.; and Tong, Y. 2020. Improving semantic segmentation via decoupled body and edge supervision. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16*, 435–452. Springer.

Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; and Yang, M.-H. 2017b. Universal style transfer via feature transforms. *Advances in neural information processing systems*, 30.

Li, Y.; Keuper, M.; Zhang, D.; and Khoreva, A. 2024. Adversarial supervision makes layout-to-image diffusion models thrive. In *The Twelfth International Conference on Learning Representations*.

Li, Y.; Zhang, D.; Keuper, M.; and Khoreva, A. 2023b. Intra-source style augmentation for improved domain generalization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 509–519.

Liu, Y.; Deng, J.; Gao, X.; Li, W.; and Duan, L. 2021. Bapa-net: Boundary adaptation and prototype alignment for cross-domain semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 8801–8811.

Luo, Y.; Zheng, L.; Guan, T.; Yu, J.; and Yang, Y. 2019. Taking a closer look at domain shift: Category-level adversaries

for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2507–2516.

Marmanis, D.; Schindler, K.; Wegner, J. D.; Galliani, S.; Datcu, M.; and Stilla, U. 2018. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135: 158–172.

Narayanan, M.; Rajendran, V.; and Kimia, B. 2021. Shape-biased domain generalization via shock graph embeddings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1315–1325.

Neuhold, G.; Ollmann, T.; Rota Bulo, S.; and Kontschieder, P. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, 4990–4999.

Ngoc, M. Ô. V.; Chen, Y.; Boutry, N.; Chazalon, J.; Carlinet, E.; Fabrizio, J.; Mallet, C.; and Géraud, T. 2021. Introducing the Boundary-Aware loss for deep image segmentation. In *British Machine Vision Conference (BMVC) 2021*.

Niu, H.; Xie, L.; Lin, J.; and Zhang, S. 2025. Exploring semantic consistency and style diversity for domain generalized semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 6245–6253.

Pan, X.; Luo, P.; Shi, J.; and Tang, X. 2018. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proceedings of the european conference on computer vision (ECCV)*, 464–479.

Peng, D.; Lei, Y.; Hayat, M.; Guo, Y.; and Li, W. 2022. Semantic-aware domain generalized segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2594–2605.

Peng, D.; Lei, Y.; Liu, L.; Zhang, P.; and Liu, J. 2021. Global and local texture randomization for synthetic-to-real semantic segmentation. *IEEE Transactions on Image Processing*, 30: 6594–6608.

Richter, S. R.; Vineet, V.; Roth, S.; and Koltun, V. 2016. Playing for data: Ground truth from computer games. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, 102–118. Springer.

Sakaridis, C.; Dai, D.; and Gool, L. V. 2019. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 7374–7383.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10765–10775.

Shelhamer, E.; Long, J.; and Darrell, T. 2014. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440.

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2636–2645.

Yue, X.; Zhang, Y.; Zhao, S.; Sangiovanni-Vincentelli, A.; Keutzer, K.; and Gong, B. 2019. Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2100–2110.

Zhang, C.; Hu, Z.; Dai, S.; He, Q.; Liu, D.; Yan, K.; and Wang, P. 2024. Boundary-Aware Contrastive Learning for Single-Source Domain Generalization in Medical Image Segmentation. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE.

Zhao, Y.; Zhong, Z.; Zhao, N.; Sebe, N.; and Lee, G. H. 2022. Style-hallucinated dual consistency learning for domain generalized semantic segmentation. In *European conference on computer vision*, 535–552. Springer.

Zhong, Z.; Zhao, Y.; Lee, G. H.; and Sebe, N. 2022. Adversarial style augmentation for domain generalized urban-scene segmentation. *Advances in neural information processing systems*, 35: 338–350.

Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; and Loy, C. C. 2022. Domain generalization: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(4): 4396–4415.