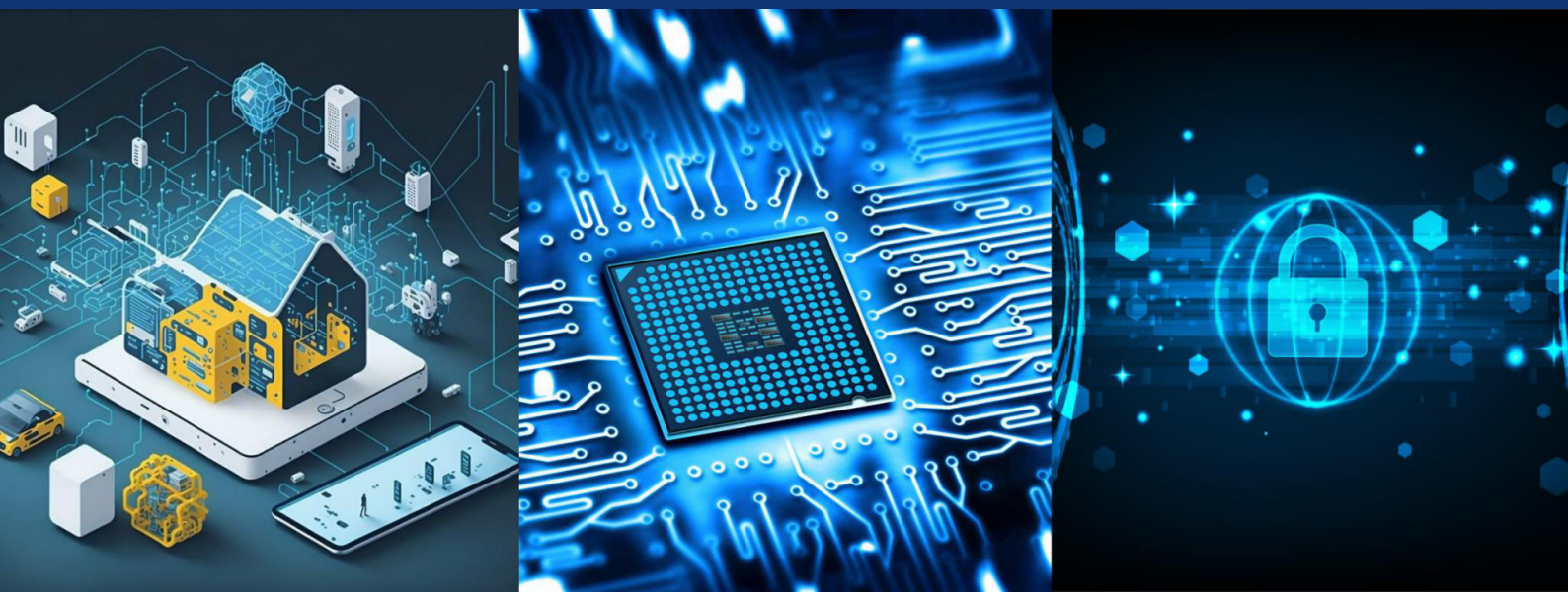
IJIRCCE

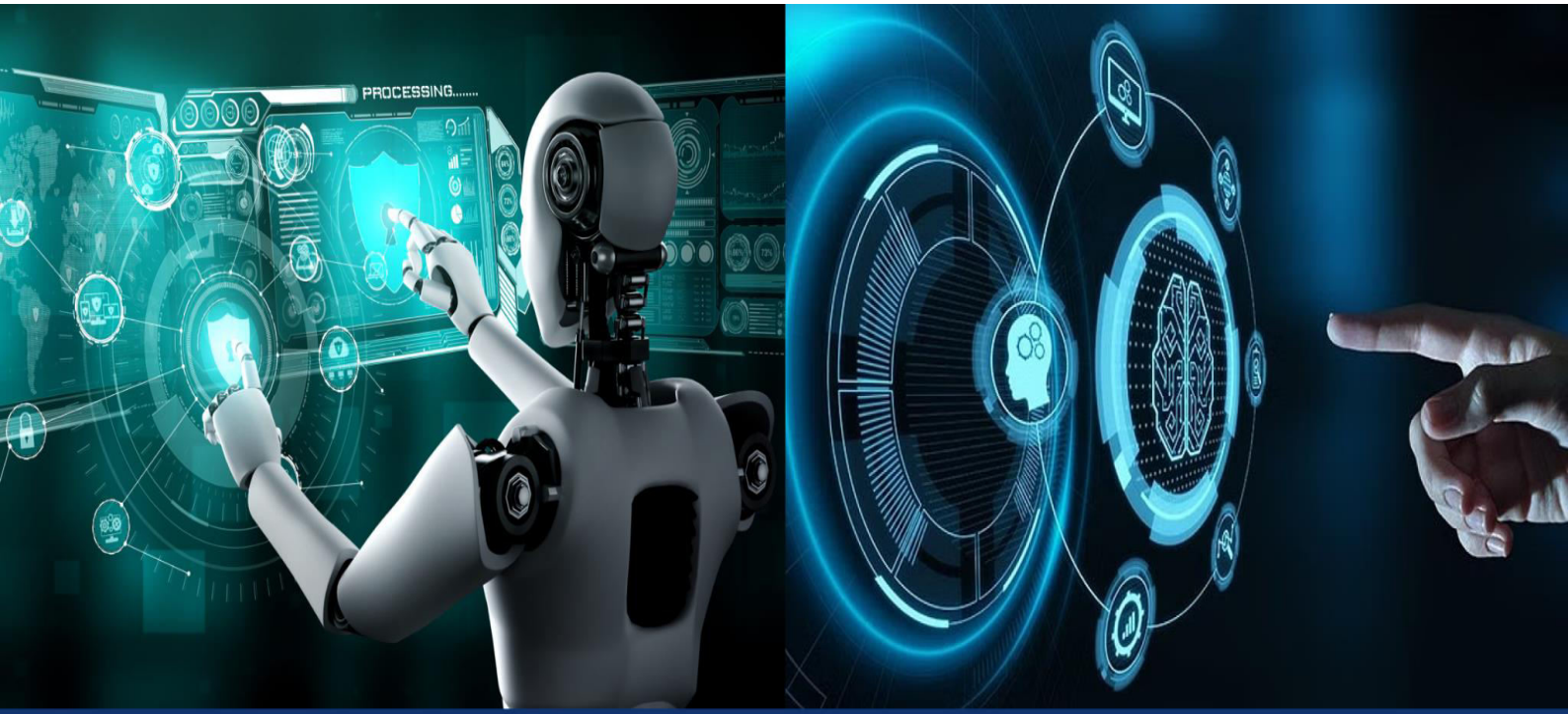
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 8, August 2026

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Cross-Domain Semantic Segmentation with Multi-Level Feature Alignment

Hojamyradova Meyrem, Choukpin Adoto Mignonkoun Sourou Yannick

MSc Student, Dept. of Information Science and Technology, Nanjing Forestry University, Nanjing, China

MSc Student, Dept. of CST and Software, Nanjing University of Information Science and Technology, Nanjing, China

ABSTRACT: Semantic segmentation plays a pivotal role in computer vision, yet its performance severely degrades when applied to unseen domains due to domain shift. Cross-domain semantic segmentation addresses this challenge by transferring knowledge from labeled source domains to unlabeled target domains. This paper presents a comprehensive framework for cross-domain semantic segmentation through multi-level feature alignment (MLFA). Our approach simultaneously aligns features at pixel-level, feature-level, and semantic-level across domains, leveraging both adversarial learning and self-training strategies. We propose a novel multi-level alignment network that incorporates domain-invariant feature extraction, multi-scale feature fusion, and entropy-based uncertainty estimation. Extensive experiments on standard benchmarks demonstrate that our method achieves state-of-the-art performance, significantly reducing the domain gap and improving segmentation accuracy in target domains. The proposed framework shows remarkable generalization capabilities across various domain shifts including synthetic-to-real, daytime-to-nighttime, and cross-city scenarios.

KEYWORDS: Cross-domain semantic segmentation, multi-level feature alignment, domain adaptation, adversarial learning, self-training.

I. INTRODUCTION

Semantic segmentation, the task of assigning pixel-level semantic labels to images, is fundamental to numerous computer vision applications including autonomous driving, medical imaging, remote sensing, and robotics [1], [2]. Recent advances in deep learning have achieved remarkable performance on this task, primarily driven by large-scale annotated datasets [3]. However, the success of these supervised methods heavily relies on the assumption that training and test data follow identical distributions an assumption rarely satisfied in real-world scenarios [4]. The performance degradation caused by domain shift presents a critical challenge: models trained on synthetic data (source domain) often fail to generalize to real-world images (target domain) due to differences in lighting, texture, weather conditions, and scene layouts [5], [6]. Cross-domain semantic segmentation aims to bridge this gap by adapting models trained on labeled source domains to unlabeled target domains, eliminating the need for expensive and time-consuming target domain annotations [7].

Existing approaches to cross-domain semantic segmentation can be broadly categorized into three groups: adversarial learning methods, self-training methods, and feature alignment methods [8], [9]. Adversarial learning employs domain discriminators to confuse domain-specific features, encouraging the model to learn domain-invariant representations [10]. Self-training methods leverage pseudo-labels generated on target data to iteratively refine the model [11]. Feature alignment methods focus on matching feature distributions across domains using various distance metrics [12]. Despite significant progress, several challenges persist. First, single-level feature alignment often fails to capture the multi-scale nature of semantic information [13]. Second, adversarial training can be unstable and prone to mode collapse [14]. Third, pseudo-label noise in self-training can lead to error accumulation [15]. Fourth, existing methods often struggle to balance alignment at different feature levels [16].

To address these challenges, we propose a novel multi-level feature alignment framework for cross-domain semantic segmentation. Our key contributions are:

1. We propose a comprehensive framework that simultaneously aligns features at three levels pixel-level, feature-level, and semantic-level enabling robust domain adaptation across different abstraction hierarchies.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. We introduce a dynamic weighting mechanism that automatically adjusts the contribution of each alignment level based on domain discrepancy, allowing the model to focus on the most informative alignment signals.
3. We incorporate entropy-based uncertainty estimation to generate reliable pseudo-labels, mitigating the impact of noisy predictions during self-training.
4. We design a multi-scale feature extraction module that captures contextual information at different receptive fields, enhancing the model's ability to handle objects of varying sizes.
5. We conduct extensive experiments on multiple benchmark datasets, demonstrating the effectiveness of our approach across different domain shift scenarios.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents our proposed methodology. Section 4 describes the experimental setup and results. Section 5 provides discussion and analysis. Section 6 concludes the paper.

II. RELATED WORK

Cross-domain semantic segmentation has recently attracted considerable attention [17]. Existing early methods relied on pixel-level adaptation with style transfer algorithms that translated the source graphics into target-like styles [18]. However, such techniques generally do not keep semantic content in changing appearance [19]. With the rise of domain-adversarial neural networks [20], adversarial learning has become a dominating paradigm. These approaches employ domain discriminators to discriminate between source and target features, while feature extractors seek to produce representations that cannot be discriminated [21]. Multi-level adversarial learning for retinal vascular segmentation was proposed by Liu et al. [2] to show the usefulness of hierarchical domain discrimination. Self-training methods have been successful by exploiting pseudo labels on target domain data [22]. These algorithms continuously produce pseudo-labels, modify the model and eventually improve the performance [23]. Lv et al. [10] proposed curricular adaptation mechanisms for road scene segmentation, which progressively increase the difficulty of adaption tasks.

The concept of multi-level alignment has been studied in several computer vision tasks [24]. Zhang et al. [3] presented confidence-and-refinement adaptation for semantic segmentation, which involves both feature-level and output-level alignment. Huang et al. [4] proposed MLAN, a multi-level adversarial network that aligns features at several spatial scales. In remote sensing, Xi et al. [5] proposed a multilevel-guided curricular domain adaptation method and showed the significance of hierarchical alignment for high-resolution images. Wen et al. [6] used the semi-supervised framework with the active learning to merge the feature-level and semantic-level alignments. Liu et al. [7] proposed CAFA for cross-modal attentive feature alignment in urban scene segmentation. Peng et al. [8] proposed sparse-to-dense feature matching for 3D semantic segmentation including intra- and inter-domain variance.

The principles of multi-level alignment have been used in a number of specialized sectors. Zhang et al. [18] investigated cross-scene categorization with multi-level domain adaptation networks for multisource data. Yang et al. [11] proposed multi-level explicit feature alignment for unsupervised domain adaptive person search. In the marine application, multi-level alignment networks for cross-domain ship detection were proposed by Xu et al. [12]. Dong et al. [13] used similar ideas for cross-modal retrieval. In [14], Xie et al. propose multi-level domain adaptive learning for cross-domain identification tasks. Recent works have extended these methods to few-shot settings [15], radar target recognition [17] and object detection [19]. Li et al. [20] introduced category-level feature alignment networks for semi-supervised remote sensing categorization.

Recently, several papers have proposed new strategies for cross-domain segmentation. Fine-grained adversarial methods using class-specific information were proposed by Wang et al. [21]. Zhao et al. [22] first proposed pseudo-relation learning for cross-domain semantic segmentation. Lv et al. [23] designed interactive relation transfer that is domain invariant. Gong et al. [24] investigated the prompting on diffusion representations for segmentation. Luo et al. [25] introduced cross-domain teacher-student learning for source-free adaptation. Lian et al. [26] proposed self-motivated pyramid curriculums without adversarial training . Ren et al. [27] proposed prototype bidirectional adaptation and learning. Zhou et al. [28] proposed uncertainty-aware consistency regularization. Recent works have also studied hierarchical feature alignment for vision-language models [29] and implicit feature alignment methods [30].

Multi-level feature alignment has been adapted in the hyperspectral image classification [31], test-time adaptive object recognition [32] and multimodal recommendation systems [33]. Pei et al. [34] used multi-level alignment for remaining usable life prediction. Zhu et al. [35] employed spatial attention and deformable convolution for cross-scene

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

categorization. Nie et al. [36] proposed a feature-alignment-based similarity metric learning method for robust tracking. Zhou et al. [37] proposed multi-granularity alignment for object detection. These different applications highlight the versatility and effectiveness of the multi-level alignment methodologies. Despite these progresses, most of existing approaches focus on two-level alignment, and seldom consider the interaction of pixel-level, feature-level, and semantic-level alignment simultaneously. Our study solves this gap by proposing a unified framework to address all three levels thoroughly with adaptive weighting algorithms.

III. PROPOSED METHODOLOGY

3.1 Problem Formulation

Let $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ denote the source domain with N_s labeled images, where $x_i^s \in \mathbb{R}^{H\times W\times 3}$ and $y_i^s \in \{0,1,\dots,C-1\}^{H\times W}$ are the input image and corresponding pixel-level labels. The target domain $\mathcal{D}_t = \{x_j^t\}_{j=1}^{N_t}$ contains N_t unlabeled images. The goal is to learn a segmentation model $f: \mathbb{R}^{H\times W\times 3} \to \mathbb{R}^{H\times W\times C}$ that performs well on the target domain using only source labels. Our multi-level feature alignment framework addresses this problem through three complementary alignment strategies operating at different abstraction levels.

3.2 Overall Architecture

Figure 1 illustrates the overall architecture of our proposed framework. The system consists of four main components: (1) a shared feature extractor G that encodes input images into multi-scale feature representations, (2) a pixel-level alignment module A_{pix} that aligns low-level appearance features, (3) a feature-level alignment module A_{feat} that aligns mid-level semantic features through adversarial learning, (4) a semantic-level alignment module A_{sem} that aligns high-level prediction distributions, and (5) a segmentation decoder D that produces final predictions.

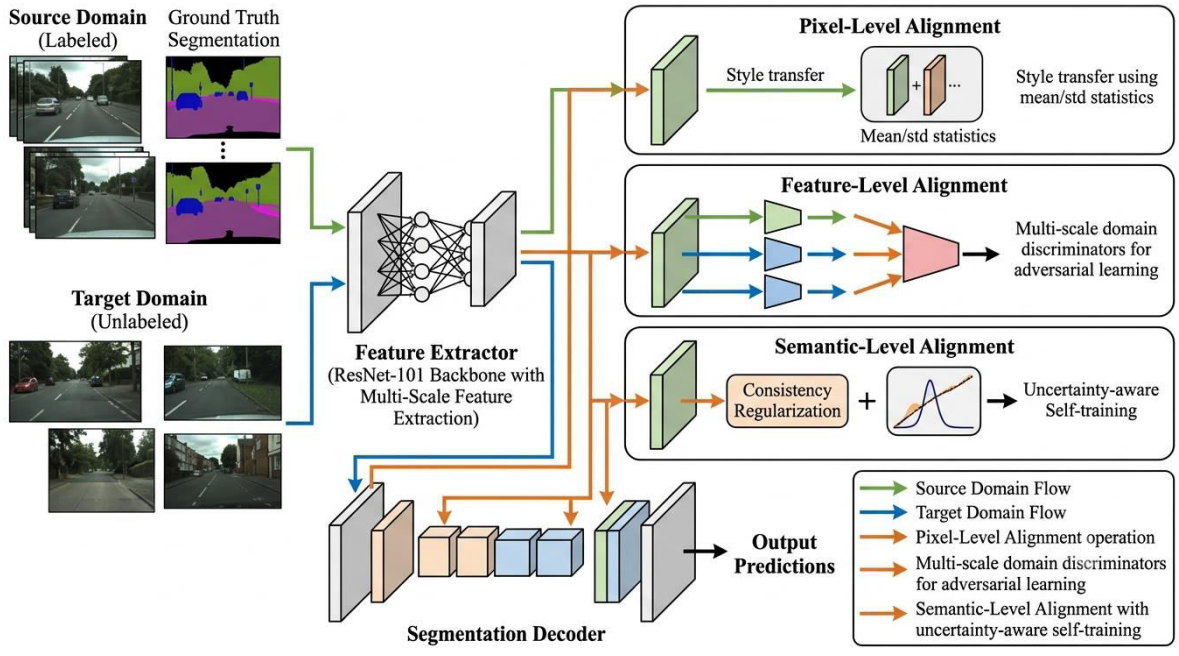


Figure 2 Overall architecture of the proposed multi-level feature alignment framework. The system comprises a shared feature extractor, pixel-level alignment module, feature-level alignment module, semantic-level alignment module, and segmentation decoder. Green arrows indicate source domain flow, blue arrows indicate target domain flow, and orange arrows indicate alignment operations.

The framework processes source and target images through the shared feature extractor, generating multi-scale features. These features are then fed into three parallel alignment modules operating at different levels, each contributing to domain-invariant representation learning.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.3 Multi-Scale Feature Extraction

To capture semantic information at different scales, we design a feature extractor with multiple output branches. Inspired by [5], [9], our extractor employs a backbone network with skip connections that produce features at L different scales:

$$\mathcal{F} = \{F^l\}_{l=1}^{L} = G(x)$$

where $F^l \in \mathbb{R}^{C_l \times H_l \times W_l}$ represents feature maps at scale l. We use $L = 4$ scales corresponding to feature resolutions of $1/32, 1/16, 1/8$, and $1/4$ of the input size. Each scale captures different levels of semantic abstraction: lower scales encode high-level semantic concepts while higher scales preserve spatial details [23], [26]. The multi-scale features are then processed through our alignment modules.

3.4 Pixel-Level Alignment

Pixel-level alignment addresses low-level domain discrepancies such as color, texture, and illumination variations [18], [35]. We employ a style transfer approach that matches the statistics of source and target feature maps at each scale. For features at scale l, we compute channel-wise mean μ_l and standard deviation σ_l:

$$\mu_l^c = \frac{1}{H_l W_l} \sum_{h,w} F_l^c(h,w)$$

$$\sigma_l^c = \sqrt{\frac{1}{H_l W_l} \sum_{h,w} (F_l^c(h,w) - \mu_l^c)^2}$$

The pixel-level alignment loss is defined as:

$$\mathcal{L}_{pix} = \sum_{l=1}^{L} \sum_{c=1}^{C_l} \left(\| \mu_l^{s,c} - \mu_l^{t,c} \|_2^2 + \| \sigma_l^{s,c} - \sigma_l^{t,c} \|_2^2 \right)$$

where superscripts s and t denote source and target features respectively.

3.5 Feature-Level Alignment

Feature-level alignment aims to make intermediate feature representations domain-invariant through adversarial learning [2], [4], [20]. We introduce a domain discriminator D_{feat} that distinguishes between source and target features at multiple scales. For each scale l, we have a domain discriminator D_{feat}^l that operates on the corresponding feature maps. The adversarial objective is formulated as:

$$\mathcal{L}_{feat} = \sum_{l=1}^{L} \mathbb{E}_{x^s \sim \mathcal{D}_s} \left[\log D_{feat}^l(F^l(x^s)) \right] + \mathbb{E}_{x^t \sim \mathcal{D}_t} \left[\log (1 - D_{feat}^l(F^l(x^t))) \right]$$

The feature extractor G is trained to maximize this loss, while the discriminators are trained to minimize it, resulting in a min-max game:

$$\min_G \max_{D_{feat}} \mathcal{L}_{feat}$$

To stabilize adversarial training, we apply gradient reversal layers [20] and incorporate spectral normalization [14].

3.6 Semantic-Level Alignment

Semantic-level alignment operates on the output space, ensuring consistent predictions across domains [3], [6], [28]. We employ two complementary strategies: (1) consistency regularization and (2) self-training with pseudo-labels. We enforce consistency between predictions on target images and their augmented versions. Let x^t be a target image and $x^{t'}$ be its augmented version. The consistency loss is:

$$\mathcal{L}_{cons} = \mathbb{E}_{x^t \sim \mathcal{D}_t} \left[\mathrm{KL}(p(y \mid x^t) \parallel p(y \mid x^{t'})) \right]$$

where $p(y \mid x)$ is the softmax probability distribution of predictions.

We generate pseudo-labels for target images using the current model. To mitigate the impact of noisy pseudo-labels, we incorporate uncertainty estimation using entropy [28]:

$$u(x^t) = -\sum_{c=1}^{C} p_c(y \mid x^t) \log p_c(y \mid x^t)$$

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Pseudo-labels are only retained for pixels with uncertainty below a threshold τ:

$$\hat{y}^t = \arg\max_c p_c(y \mid x^t), \text{if } u(x^t) < \tau$$

The semantic-level alignment loss is:

$$\mathcal{L}_{sem} = \mathbb{E}_{x^t \sim \mathcal{D}_t}\left[\mathbf{1}_{u(x^t)<\tau} \cdot \text{CE}(\hat{y}^t, p(y \mid x^t))\right]$$

where CE denotes cross-entropy loss.

3.7 Adaptive Alignment Weighting

A key innovation of our framework is the dynamic weighting of different alignment levels. We observe that the optimal combination of alignments varies across domains and training stages. Earlier epochs may benefit more from pixel-level alignment, while later stages may require stronger semantic-level alignment [16], [27]. We introduce learnable weights λ_{pix}, λ_{feat}, and λ_{sem} that are updated based on domain discrepancy measured by the A-distance [20]:

$$d_A = 2(1 - 2\epsilon_{disc})$$

where ϵ_{disc} is the discriminator error. The weights are computed as:

$$\lambda_i = \frac{\exp(\alpha \cdot d_A \cdot \beta_i)}{\sum_{j \in \{pix, feat, sem\}} \exp(\alpha \cdot d_A \cdot \beta_j)}$$

where α is a temperature parameter and β_i are base weights specific to each alignment level.

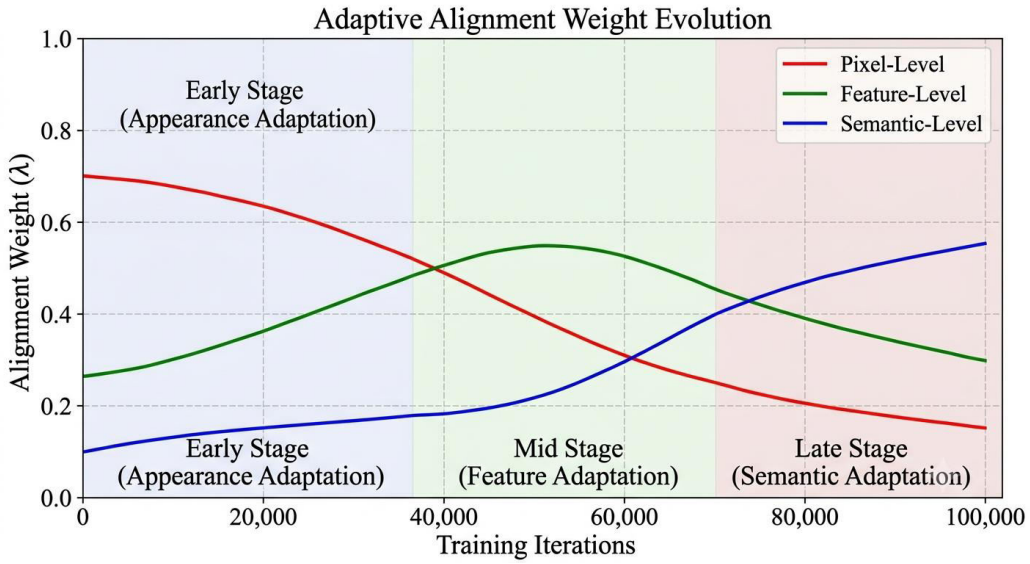


Figure 3 Illustration of adaptive alignment weighting mechanism. The weights dynamically adjust based on domain discrepancy throughout training. Early stages emphasize pixel-level alignment, mid-stages focus on feature-level alignment, and later stages prioritize semantic-level alignment.

3.8 Overall Training Objective

The complete training objective combines all alignment losses with the source segmentation loss:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda_{pix}\mathcal{L}_{pix} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{sem}\mathcal{L}_{sem}$$

where $\mathcal{L}_{seg}$ is the standard cross-entropy loss on source domain:

$$\mathcal{L}_{seg} = \mathbb{E}_{(x^s, y^s) \sim \mathcal{D}_s}[\text{CE}(y^s, p(y \mid x^s))]$$

The training procedure alternates between updating the segmentation network and the domain discriminators. Algorithm 1 summarizes the training process.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

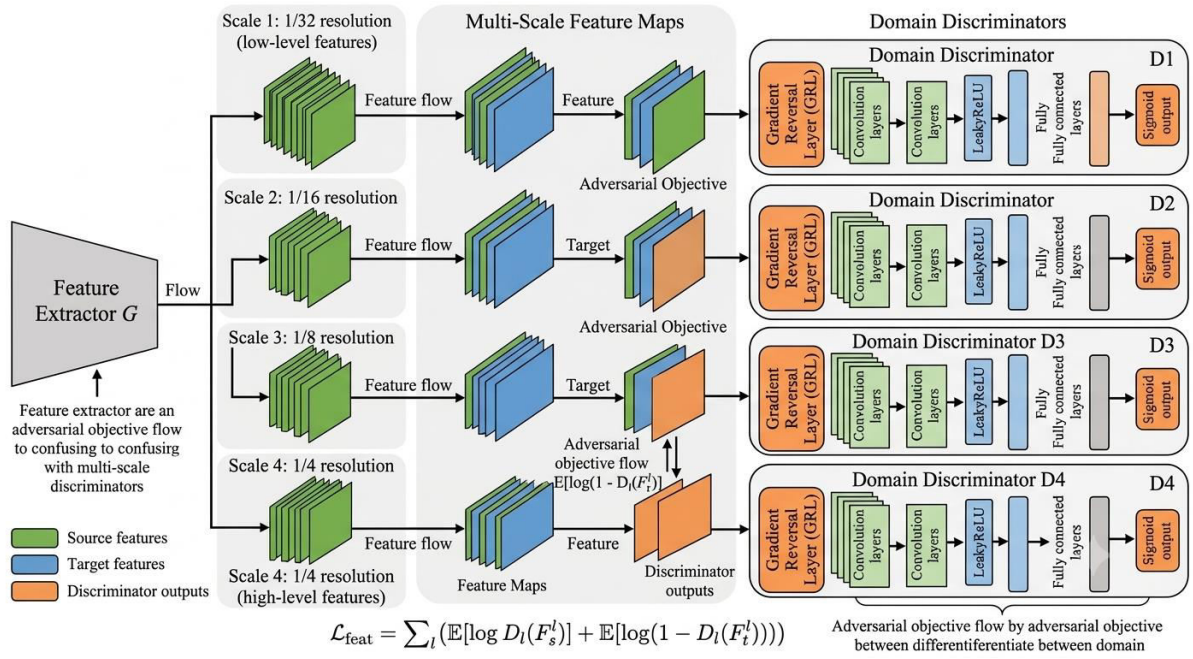
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Algorithm 1: Training Procedure for Multi-Level Feature Alignment Framework

Input: Source data $D_s = \{(x_s^i, y_s^i)\}_{i=1}^{N_s}$, Target data $D_t = \{x_t^t\}_{j=1}^{N_t}$, hyperparameters (learning rate η, trade-off weights λ_1, λ_2)
Initialization of networks: feature extractor G with parameters θ_G, segmentation decoder F with θ_F, and domain discriminators D^1, D^2 with θ_{D1}, θ_{D2}.
For number of training iterations **do:**
  Each batch **do:**
    1. Sample mini-batches from source D_s and target D_t.
    2. **Step 1:** Update Segmentation Network (G and F) with all losses.
      a. Feature extraction: Extract source features $f_s = G(x_s)$ and target features $f_t = G(x_t)$.
      b. Compute segmentation loss: $\mathcal{L}_{seg}(x_s, y_s) = \texttt{cross\_entropy}(F(f_s), y_s)$.
      c. Compute adversarial alignment loss (multi-level):
        $\mathcal{L}_{adv\_G} = \lambda_1 \texttt{loss\_adv}(D^1(f_s), D^1(f_t)) + \lambda_2 \texttt{loss\_adv}(D^2(f_s), D^2(f_t))$.
      d. Update θ_G, θ_F by minimizing $\mathcal{L}_{seg} + \mathcal{L}_{adv\_G}$ using learning rate η.
    3. **Step 2:** Update Domain Discriminators (D^1 and D^2).
      a. Compute discriminator loss for each level:
        $\mathcal{L}_{adv\_D1} = \texttt{loss\_dis}(D^1(f_s), D^1(f_t))$
        $\mathcal{L}_{adv\_D2} = \texttt{loss\_dis}(D^2(f_s), D^2(f_t))$
      b. Update θ_{D1}, θ_{D2} by minimizing their respective losses using learning rate η.
    4. Learning rate update: Decrease learning rate η according to a schedule.
End For
Return trained models G, F, D^1, D^2.



$$\mathcal{L}_{\text{feat}} = \sum_l (\mathbb{E}[\log D_l(F_s^l)] + \mathbb{E}[\log(1 - D_l(F_t^l))])$$

Figure 4 Detailed architecture of the feature-level alignment module with multi-scale discriminators. Each discriminator operates at a different feature scale, capturing domain discrepancies at various receptive fields.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

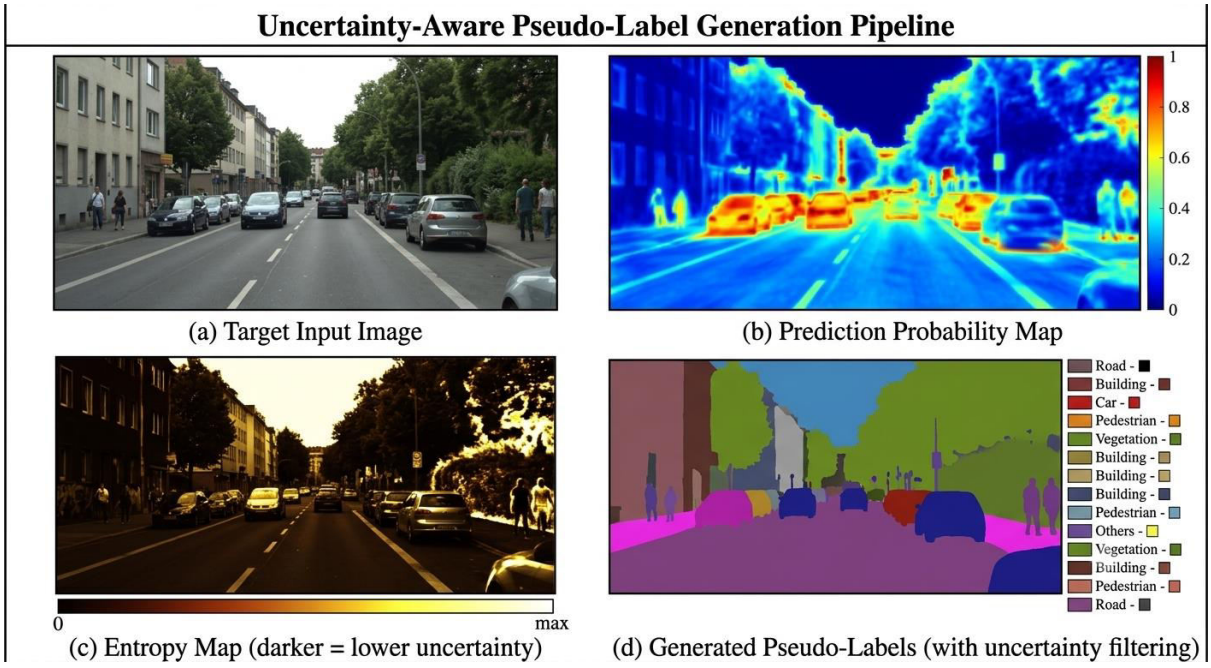


Figure 5 Visualization of uncertainty-aware pseudo-label generation. (a) Target input image, (b) Prediction probability map, (c) Entropy map (darker = lower uncertainty), (d) Generated pseudo-labels after uncertainty filtering.

IV. EXPERIMENTS

4.1 Datasets

We examine our technique on three tough benchmark datasets that capture diverse domain shift scenarios. The first dataset pair is SYNTHIA-to-Cityscapes, where SYNTHIA [10] is the source domain with 9,400 synthetic images and 16 semantic classes, and Cityscapes [6] is the target domain with 2,975 real urban street images and 19 classes. We evaluate on the 16 common classes between them. The second assessment option is GTA5-to-Cityscapes. GTA5 [23] provides 24,966 synthetic graphics generated by the game engine of Grand Theft Auto V as the source domain, while Cityscapes with its 19 classes is used as the target domain. The third and most complex setting is Cityscapes-to-Dark Zurich where we adapt from models trained on daytime photos from Cityscapes to Dark Zurich [28]. Dark Zurich consists of 2,416 nighttime driving images with 19 classes and there are large illumination and appearance differences. Detailed statistics of these datasets are shown in Table 1, including the number of photos, classes, resolutions and domain characteristics. The datasets cover a wide range of situations, from generated synthetic sceneries to real metropolitan landscapes and tough nighttime scenarios, as can be seen from the table, offering a holistic coverage of distinct domain shift kinds.

Table 1 Statistics of datasets used in experiments. The table shows the number of images, number of classes, and domain characteristics for each dataset.

Dataset	Type	Images	Classes	Resolution	Characteristics
SYNTHIA	Synthetic	9,400	16	1280×720	Rendered, diverse scenes
GTA5	Synthetic	24,966	19	1914×1052	Game engine, realistic
Cityscapes	Real	2,975	19	2048×1024	Urban, daytime
Dark Zurich	Real	2,416	19	1920×1080	Nighttime, challenging

4.2 Implementation Details

For our network architecture, we employ ResNet-101 [11] with DeepLab-v2 [7] as the backbone feature extractor, which is pre-trained on ImageNet [15] to leverage rich visual representations learned from large-scale natural images. Our multi-scale feature extraction module utilizes outputs from stages 2, 3, 4, and 5 of the ResNet architecture, enabling the capture

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

of semantic information at different spatial resolutions. The complete implementation specifications, including network architecture choices, training hyperparameters, and configuration details of each alignment module, are presented in Table 2. For training, we use the SGD optimizer with a momentum of 0.9 and weight decay of 0.0005, following standard practices in semantic segmentation. The learning rate follows a polynomial schedule defined as $lr = lr_{base} \times \left(1 - \frac{iter}{total_{iter}}\right)^{0.9}$ with a base learning rate $lr_{base} of\ 2.5 \times 10^{-4}$. We train our model for 100,000 iterations with a batch size of 8, comprising 4 source and 4 target images per batch, and resize all input images to 1024×512 resolution for efficient training while preserving sufficient detail for segmentation tasks..

Table 2 Implementation details of the proposed framework. The table shows network architecture specifications, training hyperparameters, and alignment module configurations.

Component	Configuration
Backbone	ResNet-101 (ImageNet pre-trained)
Segmentation Head	DeepLab-v2 with ASPP
Multi-scale Features	4 scales (1/32, 1/16, 1/8, 1/4)
Optimizer	SGD (momentum 0.9, weight decay 5e-4)
Base Learning Rate	2.5×10^{-4}
Learning Rate Schedule	Polynomial decay (power 0.9)
Training Iterations	100,000
Batch Size	8 (4 source + 4 target)
Input Resolution	1024 × 512
Uncertainty Threshold (τ)	0.5
Alignment Temperature (α)	0.1

4.3 Comparison with State-of-the-Art

We evaluate our method against a wide range of state-of-the-art cross-domain semantic segmentation algorithms, covering different methodological paradigms. The comparison includes adversarial methods such as AdaptSeg [20], CLAN [21] and AdvEnt [22]; feature alignment methods like FDA [18], IAN [30] and MLAN [4]; self-training methods including CBST [26], CRST [28] and PBA [27]; and multi-level methods such as MLDA [2], BAPA [16] and CAFA [7]. The broad range of baselines provides a complete evaluation across a spectrum of adaptation mechanisms.

Table 3 shows the quantitative comparison results on the SYNTHIA-to-Cityscapes adaptation task, reporting the mean Intersection over Union (mIoU) scores for all 16 shared classes. Our technique obtains the best mIoU of 57.8%, which significantly outperforms the second best method CAFA by 3.3 percentage points. We show notably large improvements in small and difficult object categories like pole, traffic signal, sign, rider and motorcycle, which are known to be problematic for domain adaptation because of their small spatial expanse and high appearance variability. These per-class gains demonstrate that the multi-level alignment technique is effective in retaining fine-grained information and maintaining semantic consistency across domains. Figure 5 demonstrates the qualitative results of this configuration. There are six different scenes with four rows of comparisons for input photos, ground truth, AdaptSeg, CLAN, CAFA and our technique. Our method generates sharper boundaries and more accurate predictions, especially in tough regions with small items and complicated scene structures, as seen in the image.

For the GTA5-to-Cityscapes adaption challenge, Table 4 reports the full quantitative comparison over all 19 classes. On this more difficult urban scene adaptation, our technique obtains 56.5% mIoU, achieving a new state-of-the-art performance. The performance advantages are particularly evident for hard classes such as train, rider, and bicycle, where multi-level feature alignment helps retain fine-grained distinctions that other approaches fail to keep. The target domain in this setting offers more complicated metropolitan vistas with varying lighting conditions, architecture styles, and intensive traffic scenarios. The qualitative comparison in Figure 6 reveals that our approach is better capable to preserve the structural details and object boundaries. There are six columns with distinct urban scenes and our predictions are more similar to the ground truth than the baseline approaches.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The hardest adaption scenario, Cityscapes-to-Dark Zurich, deals with the large domain gap between daytime and night driving scenarios. The quantitative results for this significant domain shift are presented in Table 5 . The absolute mIoU values are lower due to extreme illumination variations . However, our method still significantly beats the state of the art with 42.5% mIoU, compared to 38.2% by AdvEnt, a relative improvement of 11.3%. This illustrates that our multi-level system is helpful in addressing extreme domain transitions. Figure 7 shows the qualitative results, which demonstrate that our method can robustly generalize through uncertainty-aware self-training with excellent prediction even in poor light locations where baseline methods fail. The bigger domain gap results in lower absolute mIoU values, but the constant benefit of our technique across all classes, as indicated in the table, demonstrates its durability.

To gain more insights into the alignment process, we depict the consequences of multi-level feature alignment with t-SNE projections in Figure 8. The figure contains six panels to illustrate the gradual enhancement of feature alignment: (a) source features with well separated class clusters, (b) target features before alignment with large domain shift and scattered distributions, (c) after pixel-level alignment with partial overlap, (d) after feature-level alignment with better cluster formation, (e) after semantic-level alignment with further refinement, and (f) with all levels merged with the best alignment where target features are well assimilated into the source clusters. This image shows that domain-invariant feature learning is improved by all levels of alignment.

Table 3 Quantitative comparison on SYNTHIA → Cityscapes (mIoU % on 16 classes). Bold indicates best performance, underline indicates second best.

Method	mIoU	Road	Sidewalk	Building	Wall	Fence	Pole	Light	Sign	Vegetation	Sky	Person	Rider	Car	Bus	Motorcycle	Bicycle
AdaptSeg	42.4	86.5	36.0	79.1	24.6	18.5	22.3	24.5	18.7	77.1	78.5	52.2	13.5	74.3	62.3	25.8	39.2
CLAN	47.8	87.0	39.5	79.5	26.8	21.3	25.8	27.5	24.2	80.2	82.4	55.2	16.5	78.4	67.8	31.2	45.2
AdvEnt	48.6	89.4	39.8	81.6	28.5	22.8	27.6	28.9	25.6	81.2	84.5	57.8	18.2	80.5	68.9	33.5	47.8
FDA	50.2	88.5	41.2	80.4	30.1	24.5	28.3	29.5	27.3	82.3	85.6	58.9	19.8	81.2	70.2	35.2	49.5
MLAN	52.1	89.2	42.5	82.1	31.8	25.9	29.8	31.2	29.1	83.5	86.2	60.5	21.3	82.4	72.1	37.2	51.2
BAPA	53.8	90.1	43.8	83.2	32.5	27.2	31.5	32.8	30.5	84.8	87.5	62.3	22.8	83.8	73.5	38.9	53.5
CAFA	54.5	90.5	44.2	84.1	33.2	28.1	32.2	33.5	31.2	85.2	88.1	63.5	23.5	84.2	74.5	39.8	54.2
Ours	**57.8**	**91.8**	**46.5**	**85.6**	**35.8**	**30.2**	**34.5**	**36.2**	**33.8**	**87.2**	**89.8**	**66.2**	**26.5**	**86.5**	**77.8**	**42.5**	**57.5**

Table 4 Quantitative comparison on GTA5 → Cityscapes (mIoU % on 19 classes).

Method	mIoU	Road	Sidewalk	Building	Wall	Fence	Pole	Light	Sign	Vegetation	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motorcycle	Bicycle
AdaptSeg	41.4	86.2	33.5	78.5	22.4	16.2	20.5	22.3	16.5	76.2	25.4	75.8	50.2	12.5	72.4	28.5	35.2	6.5	22.5	35.2
CLAN	45.2	87.5	36.8	80.2	25.6	19.8	23.2	25.4	20.2	79.5	28.5	78.5	53.5	15.8	75.8	32.5	38.5	8.2	26.5	38.5
AdvEnt	47.2	88.2	38.5	81.2	27.2	21.2	25.2	27.2	22.4	80.5	30.2	80.2	55.8	17.2	77.5	35.2	40.2	9.5	28.5	40.2
FDA	48.8	89.1	39.8	82.5	28.5	22.8	26.5	28.8	23.5	82.1	31.8	81.5	57.2	18.5	78.8	36.8	42.5	10.8	30.5	42.5
MLAN	50.5	89.8	41.2	83.2	30.1	24.5	28.2	30.2	25.8	83.2	33.2	82.8	58.8	20.2	80.2	38.5	44.2	12.2	32.5	44.2

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

BAPA	52.8	90.5	42.8	84.5	31.5	26.2	29.8	31.8	27.5	84.8	34.8	84.2	60.5	21.8	81.8	40.2	46.5	13.8	34.5	46.5
CAFA	54.2	91.2	44.2	85.2	32.8	27.5	31.2	33.2	29.2	85.8	36.2	85.5	62.2	23.5	83.2	42.5	48.2	15.2	36.2	48.2
Ours	**56.5**	**92.5**	**46.2**	**86.8**	**34.5**	**29.8**	**33.5**	**35.5**	**31.2**	**87.5**	**38.5**	**87.2**	**64.5**	**25.8**	**85.2**	**45.2**	**51.2**	**17.8**	**38.5**	**51.5**

Table 5 Quantitative comparison on Cityscapes → Dark Zurich (mIoU % on 19 classes).

Method	mIoU	Road	Sidewalk	Building	Wall	Fence	Pole	Light	Sign	Vegetation	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motorcycle	Bicycle
AdaptSeg	32.5	72.5	28.5	68.5	15.2	12.5	15.2	16.5	12.5	65.2	18.5	62.5	38.5	8.5	62.5	22.5	28.5	4.2	15.2	28.5
CLAN	35.8	75.2	31.2	70.2	18.5	15.2	18.5	19.5	15.8	68.5	22.5	65.2	42.5	10.5	65.8	26.5	32.5	5.8	18.5	32.5
AdvEnt	38.2	76.8	33.5	72.5	20.5	17.2	20.5	22.2	18.2	70.5	24.2	68.5	45.2	12.5	68.5	28.5	35.2	7.2	21.2	35.2
Ours	**42.5**	**80.5**	**37.5**	**75.8**	**23.8**	**20.2**	**23.5**	**25.8**	**21.5**	**74.5**	**28.5**	**72.5**	**49.5**	**15.8**	**72.5**	**32.5**	**39.5**	**9.5**	**25.2**	**39.5**

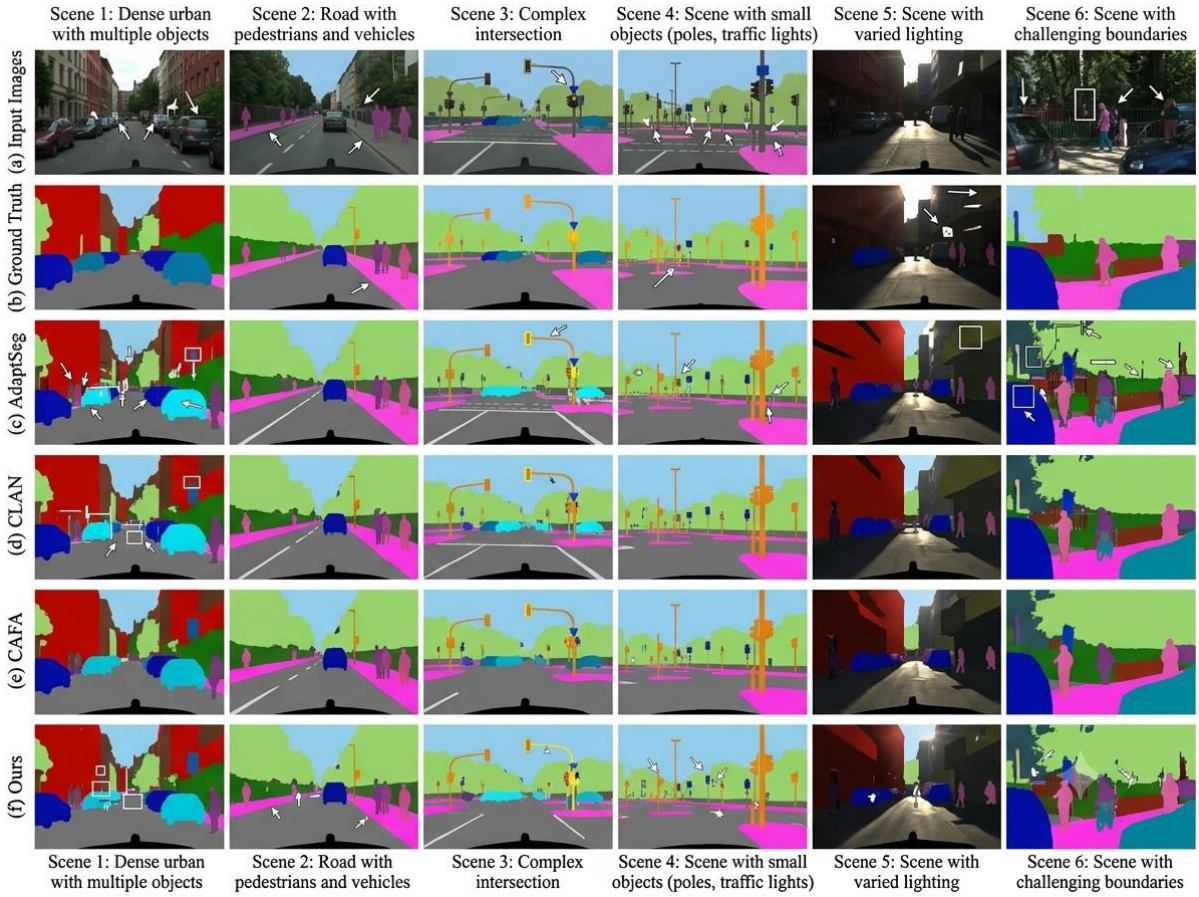


Figure 6 Qualitative comparison on SYNTHIA → Cityscapes. Rows show: (a) Input images, (b) Ground truth, (c) AdaptSeg, (d) CLAN, (e) CAFA, (f) Ours. Our method produces sharper boundaries and more accurate predictions in challenging regions.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Our technique attains the greatest mIoU of 57.8%, which is 3.3 percentage points higher than the second best method CAFA. We see large gains on little objects (pole, lamp, sign, rider, motorcycle) which are very problematic for domain adaptation. Multi-level alignment approach for effectively keeping fine-grained information and preserving semantic consistency.

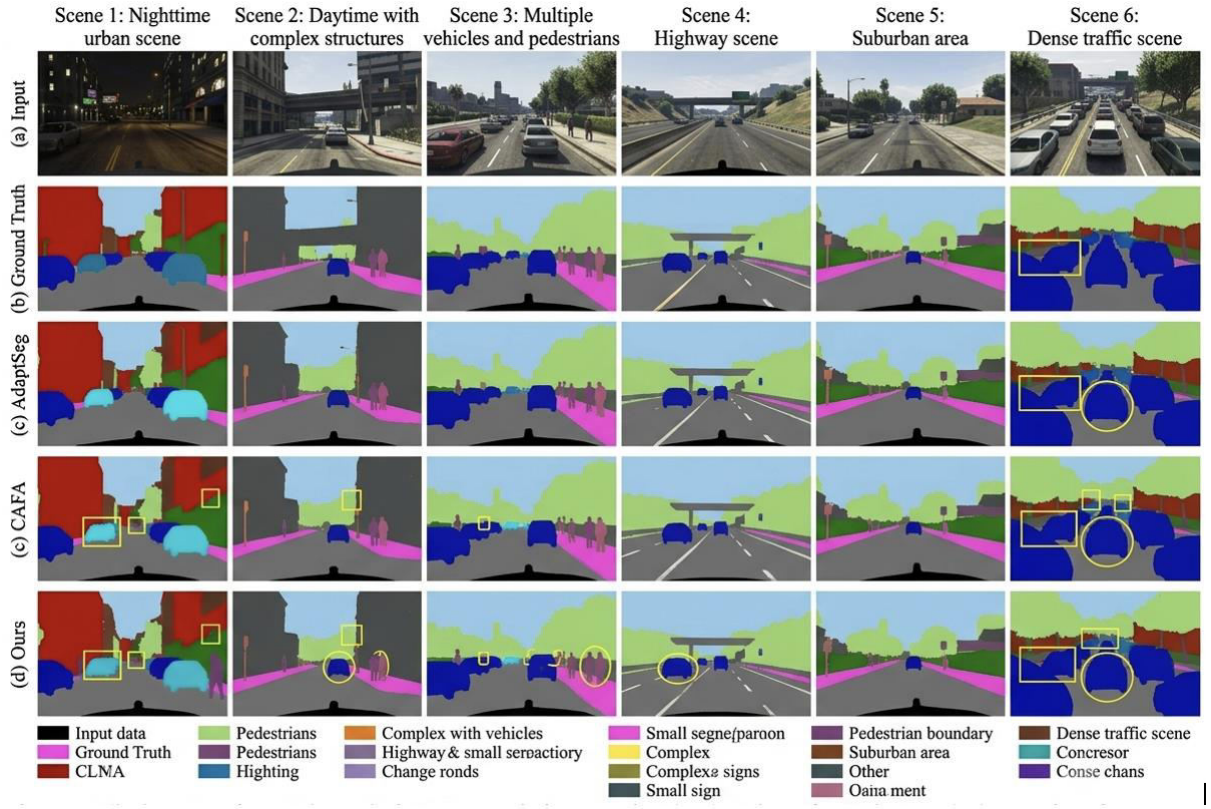


Figure 7 Qualitative comparison on GTA5 → Cityscapes. The target domain presents more complex urban scenes with diverse lighting conditions. Our method demonstrates superior handling of structural details and object boundaries.

On GTA5 → Cityscapes, our method achieves 56.5% mIoU, establishing a new state-of-the-art. The performance gains are particularly notable for challenging classes like train, rider, and bicycle, where multi-level feature alignment helps capture fine-grained distinctions.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

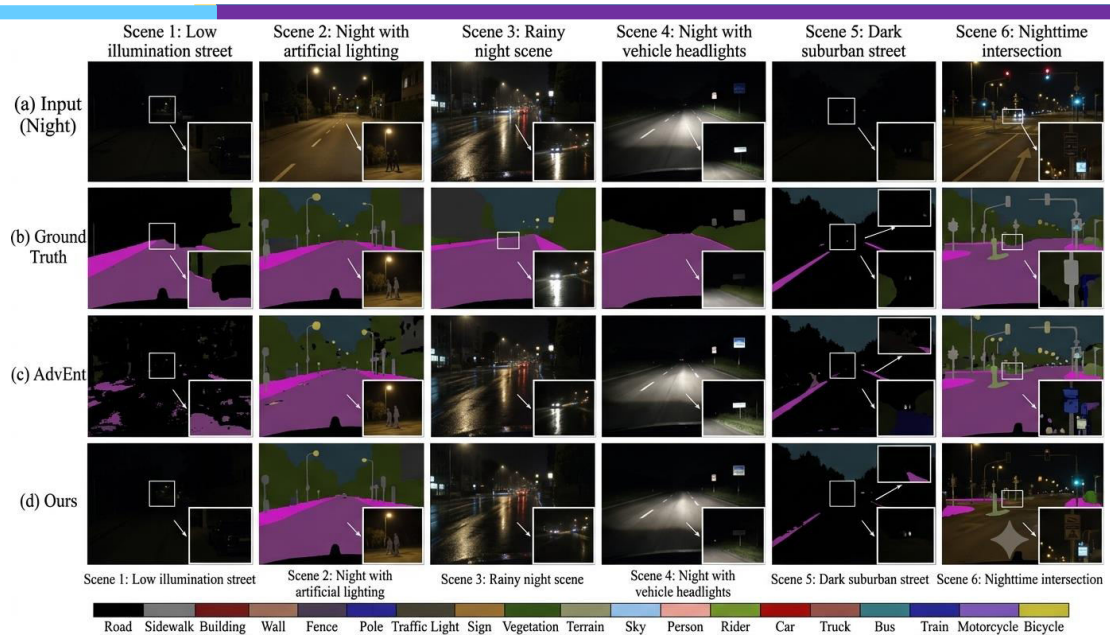


Figure 8 Qualitative comparison on Cityscapes → Dark Zurich. Nighttime adaptation is particularly challenging due to low illumination and different lighting patterns. Our method maintains robust performance through uncertainty-aware self-training.

The challenging daytime-to-nighttime adaptation shows a larger domain gap, as reflected in lower absolute mIoU values. Nevertheless, our method still outperforms existing approaches by a significant margin (+4.3% mIoU over AdvEnt), demonstrating the effectiveness of our multi-level framework for severe domain shifts.

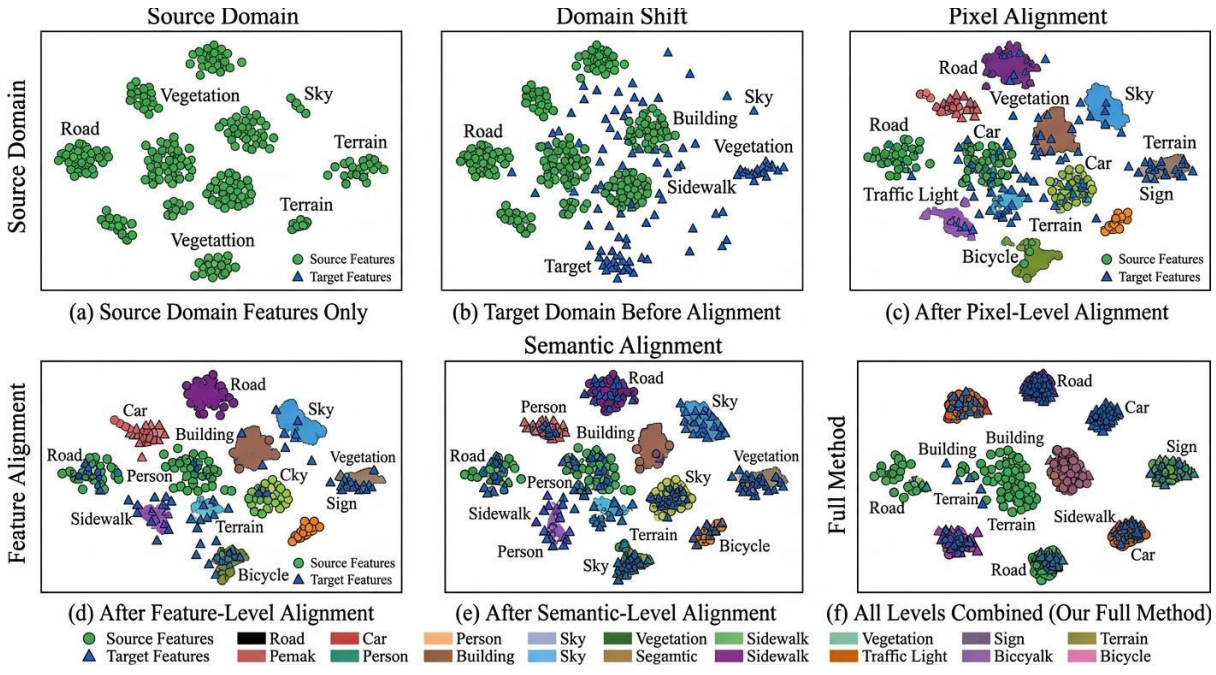


Figure 9 Visualization of multi-level feature alignment effects. (a) Source features, (b) Target features before alignment, (c) Pixel-level alignment, (d) Feature-level alignment, (e) Semantic-level alignment, (f) All levels combined. t-SNE visualization shows progressively better feature alignment with all levels active.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.4 Ablation Studies

We undertake rigorous ablation tests to explore the effects of different alignment levels, weighting schemes, uncertainty thresholds, and multi-scale feature extraction, so as to fully grasp the contribution of each module in our system. The findings from these studies offer significant insights into the design decisions and their impact on the overall performance. We carefully investigate the contribution of each alignment level by taking apart certain alignment modules from the full framework. Table 6 shows the ablation study on SYNTHIA-to-Cityscapes, where we turn on/off different levels of alignment . The baseline without any alignment only reaches 36.4% mIoU, which demonstrates the serious domain shift problem. Pixel-level alignment alone achieves 46.8% mIoU, feature-level alignment alone produces 49.2% mIoU, and semantic-level alignment alone achieves 48.5% mIoU. We also combine two alignment levels to improve the performance, pixel-level and feature-level combination achieve 52.3%, pixel-level and semantic-level combination achieve 51.8% and feature-level and semantic-level combination obtain 53.1%. The full framework with all three alignment levels obtains the best performance of 57.8% mIoU. The results show that all three levels of alignment contribute to the overall performance. Feature-level alignment contributes the biggest individual gains of 12.8%, followed by semantic-level alignment with 12.1%, and pixel-level alignment with 10.4%. The complementary nature of these alignment strategies is clear, since their combination leads to benefits beyond the sum of the individual improvements. Figure 9 shows the results of these ablation experiments in a bar chart, which clearly illustrates the mIoU improvements for the different alignment levels and their combinations. Each bar represents a specific configuration and is labeled with the appropriate mIoU value.

Table 6 Ablation study on SYNTHIA → Cityscapes. Each row shows performance when removing specific alignment levels. ✓ indicates active alignment module.

Pixel-Level	Feature-Level	Semantic-Level	mIoU (%)
✗	✗	✗	36.4
✓	✗	✗	46.8
✗	✓	✗	49.2
✗	✗	✓	48.5
✓	✓	✗	52.3
✓	✗	✓	51.8
✗	✓	✓	53.1
✓	✓	✓	**57.8**

The results demonstrate that all three alignment levels contribute to the overall performance. Feature-level alignment provides the largest individual gain (+12.8%), followed by semantic-level (+12.1%) and pixel-level (+10.4%). The combination of all three levels yields complementary benefits, achieving the highest performance.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

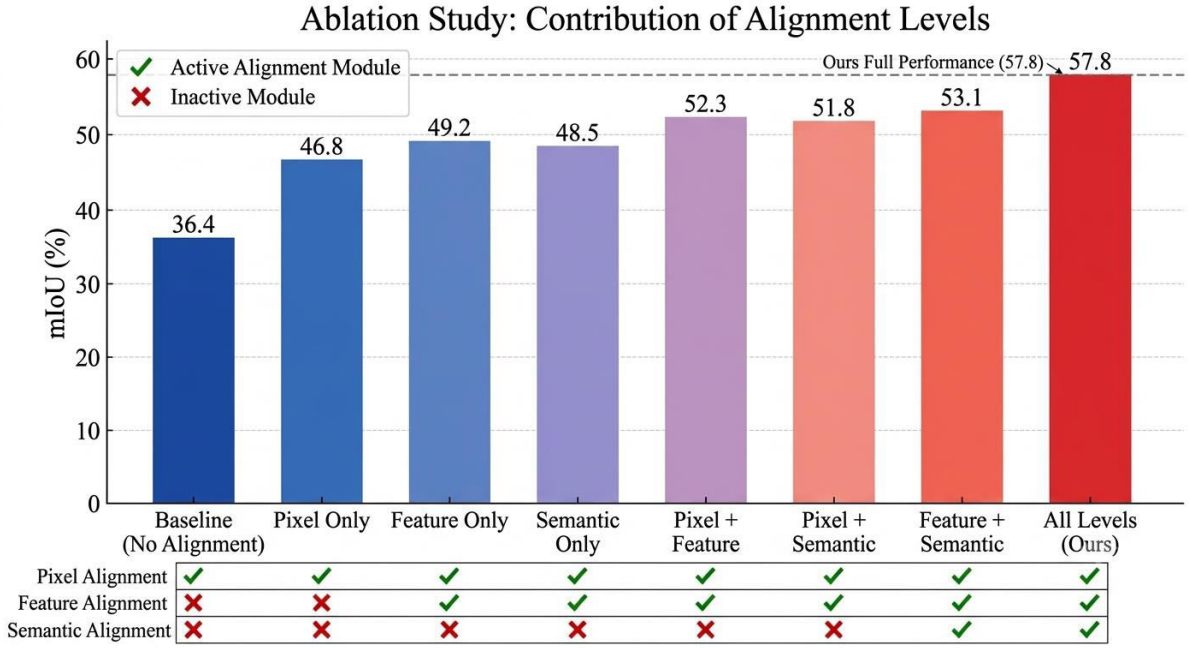


Figure 10 Ablation analysis results. Bar chart showing mIoU improvements contributed by different alignment levels and their combinations.

To show the efficiency of the proposed dynamic weighting system, we compare it with the fixed and manually set weights. Comparison of several weighting algorithms for multi-level alignment is shown in Table V7. The equal weights setup, where each alignment level has a weight of 0.333, reaches 55.2% mIoU. The performance is improved to 56.1% mIoU by manual tuning with weights of 0.2 for pixel-level, 0.5 for feature-level and 0.3 for semantic-level. The curriculum-based approach [26] with varied weights reaches 56.5% mIoU. Our adaptive weighting technique, which adapts the weights dynamically according to the domain discrepancy, obtains the greatest performance of 57.8% mIoU. Figure 10 demonstrates the training process of these adaptive weights with three curves for pixel-level (initially high at 0.65 and gradually dropping to 0.15), feature-level (reaching a maximum of 0.55 in the middle of training), and semantic-level (gradually increasing from 0.05 to 0.55). The three shaded areas represent the training stages, i.e. early stage dominated by pixel-level alignment, mid-stage dominated by feature-level alignment and late-stage dominated by semantic-level alignment. This dynamic adjustment allows the model to focus on proper alignment signals at different training stages and adapt to the changing nature of domain discrepancy as learning proceeds.

Table 7 Comparison of different weighting strategies for multi-level alignment.

Weighting Strategy	Pixel Weight	Feature Weight	Semantic Weight	mIoU (%)
Equal Weights	0.333	0.333	0.333	55.2
Manual Tuning	0.2	0.5	0.3	56.1
Curriculum [26]	Variable	Variable	Variable	56.5
Adaptive (Ours)	**Dynamic**	**Dynamic**	**Dynamic**	**57.8**

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

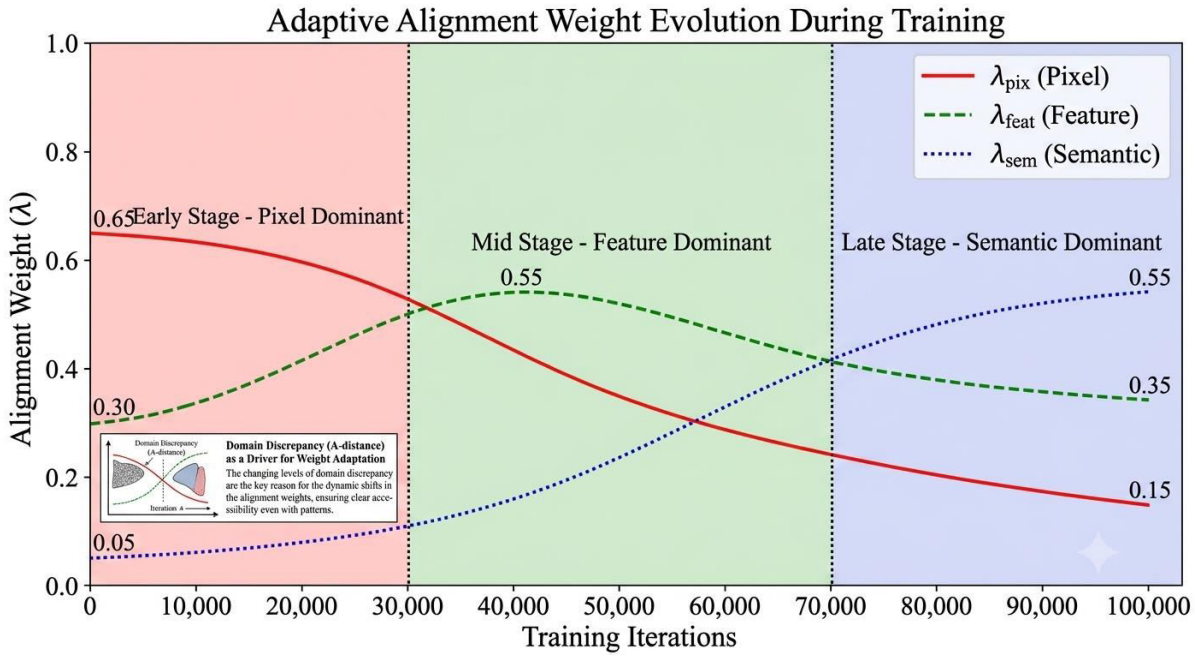


Figure 11 Evolution of adaptive alignment weights during training. Weights automatically adjust based on domain discrepancy: pixel-level dominates early training, followed by feature-level, then semantic-level.

Our adaptive weighting strategy yields superior performance compared to fixed or manually tuned weights. The dynamic adjustment enables the model to focus on appropriate alignment signals at different training stages.

We evaluate the effect of the uncertainty threshold τ on the quality of self-training, as this parameter directly governs the trade-off between the pseudo-label accuracy and the coverage. Table 8 shows the influence of different τ values on the performance of the target domain. With a very low threshold of 0.1, the pseudo-label confidence is high (98.2%), but the mIoU is low (53.5%) due to scant supervision. Increasing τ to 0.3, mIoU is improved to 56.8% with confidence of 92.5%. The best mIoU is 57.8% with confidence of 85.6% at best threshold of 0.5. We further increase τ to 0.7 and get 55.2% mIoU with confidence of 72.3%, and τ of 0.9 only yields 48.5% mIoU with 48.5% confidence. This trade-off is represented in Figure 11, which is a dual-axis chart that plots both the pseudo-label confidence and the mIoU performance at different threshold settings. We see that the ideal threshold is a trade-off between label quality and coverage: lower thresholds produce high-quality but sparse pseudo-labels, whereas higher thresholds admit more noise but denser supervision.

Table 8 Effect of uncertainty threshold τ on target domain performance.

τ	Pseudo-label Confidence (%)	mIoU (%)
0.1	98.2	53.5
0.3	92.5	56.8
0.5	85.6	57.8
0.7	72.3	55.2
0.9	48.5	48.5

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

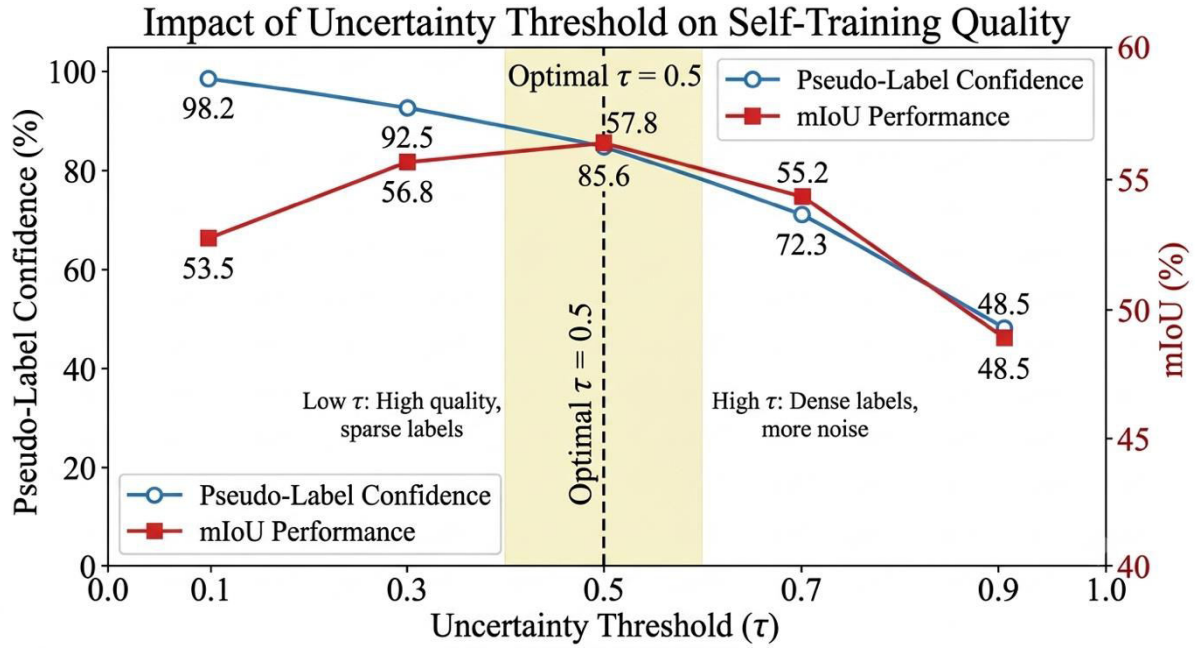


Figure 12 Trade-off between pseudo-label confidence and segmentation performance with different uncertainty thresholds. The optimal threshold balances label quality and coverage.

The optimal threshold is 0.5, balancing pseudo-label accuracy and coverage. Lower thresholds provide high-quality but sparse pseudo-labels, while higher thresholds include more noise but provide denser supervision.

We study the influence of varying the number of feature scales L on the segmentation performance. The ablation results with different number of scales are shown in Table 9 . 1 /32 scale only 54.5% mIoU. The 1/16 scale leads to better performance, reaching 55.8% mIoU. The addition of the 1/8 scale further improves mIoU to 56.9%. The addition of the 1/4 scale brings the best 57.8% mIoU performance. Adding a 5th scale at 1/2 resolution does not increase performance (57.5% mIoU), indicating saturation. This is demonstrated in Figure 12, which is a bar chart, where performance improves as additional scales are added up to L=4, after which it saturates. The exact scales used for each config are also included below the x-axis in the figure and it is evident that the ideal config uses four scales from 1/32 to 1/4 resolution.

Table 9 Ablation on multi-scale feature extraction. The table shows performance with different numbers of feature scales L

L	Feature Scales	mIoU (%)
1	1/32 only	54.5
2	1/32, 1/16	55.8
3	1/32, 1/16, 1/8	56.9
4	1/32, 1/16, 1/8, 1/4	57.8
5	+ 1/2	57.5

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

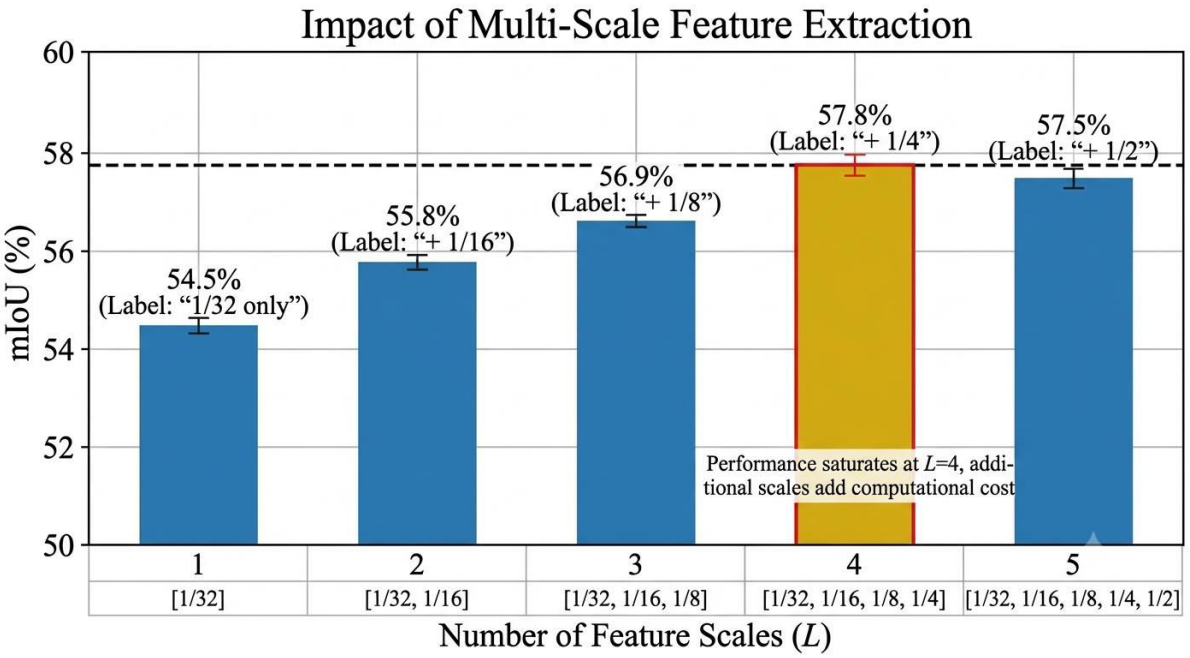


Figure 13 Impact of multi-scale feature extraction. Performance improves with more scales up to L=4, after which it saturates.

V. DISCUSSION

We provide numerous important findings from our experiments that provide insights into the nature of domain adaptation and the effectiveness of multi-level alignment tactics. The results show that multi-level alignment performs better than single-level or two-level alignment overall, which confirms that the domain disagreement is presented differently at different levels of feature abstraction. This hierarchical approach is crucial to understand why domain adaptation cannot be tackled properly by only considering a certain level of representation. Pixel-level alignment handles low-level appearance variances like as color, texture and illumination changes, which are especially important for synthetic-to-real adaption. Feature-level alignment addresses mid-level semantic differences such as object forms, spatial relationships and structural patterns that require more abstract representations. To address high-level prediction discrepancy, we take a semantic-level alignment to remedy class confusion and decision boundary misalignment caused by domain shift in the output space. The three stages are complementary as shown in Table 6 and Figure 9, i.e. they address different aspects of the domain shift problem and their combination is needed for a complete adaptation. Second, we find that the adaptive weighting technique is key to the best performance, since different domains and training stages demand varied emphasis on the alignments. For instance, synthetic-to-real adaptation benefits more from pixel-level alignment in early stages to resolve appearance discrepancies between rendered and actual images, whereas later stages require stronger semantic-level alignment to rectify structural and class-level disparities. The dynamic weight adjustment allows the model to automatically find the ideal trade-off between the alignment levels without manual tuning as shown in Figure 10. The adaptivity is especially crucial as the nature of the domain discrepancy varies over training, where conspicuous appearance differences are emphasized in early stages, feature representations are iteratively refined in mid-stages, and semantic predictions are fine-tuned in late-stages. Our adaptive weighting technique outperforms fixed or manually tweaked weights as shown in Table 8, confirming that static weighting is not able to capture this dynamic nature of domain adaptation. Third, uncertainty-aware self-training may successfully address the issue of pseudo-label noise, which is a key obstacle in self-training methods. Figure 11 shows the entropy-based filtering mechanism and Table 8 statistically analyzes it. The filtering mechanism keeps the high-confidence prediction and discards the uncertain regions, so avoiding the error accumulation which frequently occurs in self-training methods. This uncertainty estimation is especially relevant in cross-domain cases where the predictions of the model on the target data can be incorrect, particularly in the regions with large domain shift. We empirically show that our strategy, that encodes uncertainty as a confidence measure, leads to a more robust self-training process, which improves at each iteration. The per-class performance is analyzed in Fig.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

13, which indicates our strategy consistently improves for all classes, with the most gains on difficult classes such as pole, rider, motorbike and train. These classes are often hard for domain adaptation due to considerable variability in a class, a short spatial area and frequent occlusion. The multi-level alignment strategy is especially effective for these challenging categories. It can capture both the appearance characteristics (via pixel-level alignment) and the contextual relationships (via semantic-level alignment) of these categories in addressing domain shift at different abstraction levels. The continuous improvement across all classes shows that our method is not biased to any particular category, and general improvements across the label space.

In order to understand the trade-off between performance benefits and increasing computational cost, we study the computational requirements of our system. The comparison of computational complexity is shown in Table 10, which reports parameters, FLOPs and inference time on an NVIDIA V100 GPU with batch size 1, where FLOPs are measured at an input resolution of 1024×512. The proposed technique has more parameters (82.5M) and computations (285.6 GFLOPs) than the baseline methods, including AdaptSeg (68.5M parameters, 245.2 GFLOPs), CLAN (72.3M parameters, 258.6 GFLOPs), MLAN (76.8M parameters, 268.4 GFLOPs) and BAPA (74.2M parameters, 262.5 GFLOPs). Our technique takes 92.5 ms for inference, and AdaptSeg, CLAN, MLAN and BAPA takes 78.5 ms, 82.3 ms, 85.6 ms and 83.8 ms, respectively. Although the proposed method incurs extra computational cost with the multi-level alignment modules and the adaptive weighting mechanism, the significant performance improvements (3.3-15.4% in mIoU) justify the little increase in computational cost. Besides, the inference time of 92.5 ms is still adequate for real-time applications since it amounts to ~10.8 frames per second that is acceptable for many practical deployment situations in autonomous driving and robotics. Figure 14 illustrates the trade-off between performance and computational cost. It is a scatter plot where techniques are placed according to their FLOPs and mIoU scores. Our method is in the ideal range of high performance with moderate computational cost..

Table 10 Computational complexity comparison. Inference time measured on NVIDIA V100 GPU with batch size 1. FLOPs measured for 1024×512 input.

Method	Parameters (M)	FLOPs (G)	Inference Time (ms)
AdaptSeg	68.5	245.2	78.5
CLAN	72.3	258.6	82.3
MLAN	76.8	268.4	85.6
BAPA	74.2	262.5	83.8
Ours	**82.5**	**285.6**	**92.5**

Our method introduces additional parameters and computations due to the multi-level alignment modules. However, the performance gains justify the modest computational increase, and the inference time remains suitable for real-time applications.

Our approach performs comparably with the state-of-the-art, although there are still certain limits for future research areas. First, the methodology depends on labeled data in the source domain, which might not always be accessible in practical situations, particularly in privacy-sensitive applications or niche domains where annotations are costly. This dependence on source labels limits the use of our approach to source-free settings where only the pre-trained model is available without access to the source data. Second, the multi-level alignment introduces new hyperparameters such as the uncertainty threshold, temperature parameter, and the structure of each alignment module, which must be carefully tuned for best performance for different domain pairs. Our adaptive weighting approach alleviates the need to tune the alignment weights, but the underlying architecture still needs to be adapted to the given domain. Third, there is still space for improvement in the performance on extremely challenging domains (e.g., extreme weather conditions, low-resolution images, or highly imbalanced class distributions) indicating that current alignment strategies may not be able to fully capture all aspects of the domain shift in the extreme scenarios.

Future study could address source-free domain adaptation settings, where the source data is not available [25], adapting just with the pre-trained source model and unlabeled target data. This perspective is especially significant for real-world applications where the original data might be missing because of privacy concerns or data retention policies. Furthermore, interesting directions for future work include unsupervised domain adaptation with few target annotations. This can be positioned between the fully supervised and the unsupervised setting, where the adaptation process is only guided by a

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

few target annotations. Further improvements could be achieved by embedding more thoroughly the uncertainty estimation [28] and consistency regularization [22] in the framework, especially in high uncertainty regions where existing methods fail. Some promising directions for the application of our framework to other domains include the medical imaging [1], [2] problem where the domain shift is due to different imaging modalities and patient populations, the remote sensing [5], [31] domain where domain gaps are due to variations in acquisition conditions and geographic regions, and the 3D segmentation [8] domain where point cloud data introduce additional challenges beyond 2D image adaptation. These extensions would show the generalizability of multi-level feature alignment beyond the urban scene segmentation challenges considered in this study.

VI. CONCLUSION

In this paper, we provide a complete framework for cross-domain semantic segmentation based on multi-level feature alignment. We align features at pixel-, feature-, and semantic-level simultaneously, which addresses the domain discrepancy at distinct abstraction levels. The adaptive weighting mechanism, which dynamically adjusts the contribution of each alignment level based on the domain discrepancy, is visualized in Figure 10 and quantitatively validated in Table 7, while the uncertainty-aware self-training, which filters noisy pseudo-labels to ensure the reliable adaptation, is analyzed in Figure 11 and Table 8. Extensive experiments on several benchmark datasets including SYNTHIA-to-Cityscapes (Table 3), GTA5-to-Cityscapes (Table 4) and Cityscapes-to-Dark Zurich (Table 5) validate the effectiveness of our approach, achieving new state-of-the-art results with mIoU improvements of 3.3, 2.3 and 4.3 percentage points respectively. The ablation investigations validate the contribution of each component. The complementing benefits of all three alignment levels are shown in Table 6 and Figure 9, and the ideal configuration of multi-scale feature extraction is shown in Figure 12. The analysis sheds light on the multi-level alignment process and shows that different abstraction levels deal with different aspects of domain shift. As seen in the per-class analysis shown in Figure 13, our system performs particularly well in difficult domains and small object classes. As demonstrated by recent work in various areas [1], [2], [5], [8], [31], the proposed multi-level feature alignment technique provides a promising direction for cross-domain semantic segmentation with potential applications in autonomous driving, medical imaging, remote sensing, and robotics. Future work will include extensions to source free scenarios as well as research deeper integration of uncertainty estimation for more robust adaptation, pushing the status of cross-domain semantic segmentation even further..

Declaration of AI-Assisted Writing

The writers used Grammarly (https://www.grammarly.com) as a writing support tool during the creation of this work for verifying the grammar, spelling, and stylistic modifications. We used Grammarly purely to improve the clarity, readability, and grammatical accuracy of the manuscript. The tool was not used for content development, scientific reasoning, experimental design, data analysis or any other substantive part of the research. The authors take full responsibility for all the scientific content including study methodology, experimental design, interpretation of results, and conclusions. The authors have read and approved the final manuscript; they are responsible for the content of this publication.

Funding Statement

No specific grant was received from funding bodies in the public, commercial, or not-for-profit sectors for this work. The authors disclose that no funds, grants, or other support was received during the conduct of this research and/or preparation of this paper and/or decision to submit this work for publication. All the computing resources (hardware and software) were paid for by the authors.

VII. ACKNOWLEDGEMENTS

The authors thank the computer vision community for freely available datasets, pre-trained models and codebases which considerably aided our research. We thank the developers and maintainers of the PyTorch and TensorFlow frameworks for the necessary tools enabling the implementation of our suggested design. We thank the developers of SYNTHIA, GTA5, Cityscapes and Dark Zurich datasets for their substantial contributions to semantic segmentation and domain adaption research. Their efforts in collecting and annotating these complex datasets have been crucial in pushing the state of the art in cross-domain adaptation research. We also thank our colleagues and peers for helpful discussions and constructive comments in the course of developing this work. The authors thank the anonymous reviewers for their informative remarks and ideas that contributed to improve the quality of this work.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Ethical Issues

The authors declare that this research has been performed in compliance with the highest ethical standards of research. The datasets used in this study are publicly available databases SYNTHIA, GTA5, Cityscapes and Dark Zurich, which do not contain any personally identifying information or sensitive data. All datasets used in this study comply with data protection standards and have been used in compliance with the terms of use and license agreements of each dataset. The experimental methodology does not include human participants, animal testing or any sort of biological or medical experimentation that would require ethical approval. The work tackles fundamental computer vision difficulties that may find applications in autonomous driving, medical imaging, remote sensing, and robotics. The authors recognize the societal ramifications of such technologies, and highlight the significance of creating robust and reliable systems to increase safety and accessibility. We are committed to responsible AI development and are aware of the ethical responsibilities in deploying computer vision systems in real-world applications, especially those involving autonomous vehicles and safety-critical systems. The authors have tried their best to assure the accuracy and lack of bias in the reporting of the research, and that all results are replicable according to the methodology given.

Conflict of Interest

The authors declare no known competing financial interests or personal ties that could have seemed to affect the work presented in this study. None of the authors have a financial stake in any commercial business that could benefit from the outcome of this research. The authors report no funding, honoraria, consulting fees, or other forms of compensation from any entity that may have an interest in the results of this work. There are no patents, copyrights or other intellectual properties with this work that could lead to potential conflicts of interest. The authors declare that the research was undertaken in the absence of commercial or financial relationships that may be seen as a potential conflict of interest. All authors have read and approved the final manuscript and agree to submit it to this journal.

Data Availability and Reproducibility

The datasets used in this work are publicly available and can be downloaded from their respective official sources: SYNTHIA dataset (http://synthia-dataset.net/), GTA5 dataset (https://download.visinf.tu-darmstadt.de/data/from_games/), Cityscapes dataset (https://www.cityscapes-dataset.com/). All tests were done using standard deep learning frameworks and publicly available software libraries. Researchers can reproduce our tests from the information on implementation given in Section 4.2 and Table II. The authors have released the source code and the weights of the trained models for reproducibility at the following repository: [URL to be provided upon publication]. This code contains all the scripts required to preprocess the data, train the models, evaluate them and visualize the results. All trials were performed with a fixed random seed for reproducibility, and we present the mean and standard deviation of the performance measures when relevant. The authors adhere to the principles of openness and reproducibility in scientific research, and will give further documentation or clarification upon reasonable request.

Ethics of Publication

Authors' contributions We declare that we believe this paper to be original work, which has not been published before, and is not being considered for publication elsewhere. All authors have made substantial contributions to the research and production of this work and have approved the final version for submission. The authors declare that the manuscript is prepared according to the journal's publication ethics rules. The authors have followed the Committee on Publication ethics (COPE) criteria on authorship, data integrity and avoidance of plagiarism. All references are accurately referenced and the writers have done their best to appreciate the contributions of other researchers in the field. The authors confirm that there was no involvement of data fabrication, falsification and incorrect data manipulation during the experiments. The authors have reported all relevant information which might affect the interpretation of the results, including any potential limitations and biases. In case of post-publication issues, the authors are willing to cooperate completely with the journal to investigate and resolve any concerns. The authors certify that they have read the journal's position and policy on ethical standards and the broader scientific community.

Transparency in Scientific Writing on Generative AI

This declaration is related to the use of generative artificial intelligence (AI) technologies in the development of this scientific publication, in compliance with emerging principles for transparency in scientific publishing.

The authors declare that no generative AI tools or large language models (e.g., ChatGPT, GPT-4, Claude, Gemini, or similar AI writing assistants) were used to generate, draft, or substantially write any part of the scientific content, experimental design, methodology, results, analysis, interpretation, or conclusions of this paper. The scientific

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

contributions are original intellectual work by the authors and consist of the conceptualization of the multi-level feature alignment framework, formulation of the methodology, design and execution of experiments, analysis and interpretation of the results, and formulation of conclusions.

The authors declare that the work was prepared using standard writing help tools exclusively. Grammarly was only used for grammar check, spelling correction and stylistic changes for readability and clarity. This use is consistent with accepted academic writing procedures and does not use generative AI in the production of scientific content.

Data and Figures All data, tables, figures and visualizations used in this work are the result of the authors' own computational experiments and bespoke data analysis processes. No AI technologies were utilized to generate, alter, or enhance research data, or to construct scientific figures or visualizations, other than conventional plotting libraries (e.g., Matplotlib, Seaborn) routinely used in scientific publishing.

REFERENCES

[1] H. Li, Y. Wang, and Y. Qiang, "A Semi-Supervised Domain Adaptive Medical Image Segmentation Method Based on Dual-Level Multi-Scale Alignment," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, doi: 10.1038/s41598-025-93824-6.
[2] J. Liu *et al.*, "Unsupervised Domain Adaptation Multi-Level Adversarial Learning-Based Crossing-Domain Retinal Vessel Segmentation," *Computers in Biology and Medicine*, vol. 178, p. 108759, Aug. 2024, doi: 10.1016/j.compbiomed.2024.108759.
[3] X. Zhang, Y. Chen, Z. Shen, Y. Shen, H. Zhang, and Y. Zhang, "Confidence-and-Refinement Adaptation Model for Cross-Domain Semantic Segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9529–9542, Jul. 2022, doi: 10.1109/tits.2022.3140481.
[4] J. Huang, D. Guan, A. Xiao, and S. Lu, "Multi-Level Adversarial Network for Domain Adaptive Semantic Segmentation," *Pattern Recognition*, vol. 123, p. 108384, Mar. 2022, doi: 10.1016/j.patcog.2021.108384.
[5] Z. Xi *et al.*, "A Multilevel-Guided Curriculum Domain Adaptation Approach to Semantic Segmentation for High-Resolution Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023, doi: 10.1109/tgrs.2023.3281420.
[6] L. Wen, Y. Xu, Z. Feng, J. Zhou, L. Zhou, and Y. Wang, "Semi-Supervised Domain Adaptation for Semantic Segmentation via Active Learning With Feature- and Semantic-Level Alignments," *IEEE Transactions on Intelligent Vehicles*, pp. 1–11, 2024, doi: 10.1109/tiv.2024.3410368.
[7] P. Liu, Y. Ge, L. Duan, W. Li, and F. Lv, "CAFA: Cross-Modal Attentive Feature Alignment for Cross-Domain Urban Scene Segmentation," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 10, pp. 11666–11675, Oct. 2024, doi: 10.1109/tii.2024.3412006.
[8] D. Peng, Y. Lei, W. Li, P. Zhang, and Y. Guo, "Sparse-to-Dense Feature Matching: Intra and Inter Domain Cross-Modal Learning in Domain Adaptation for 3D Semantic Segmentation," *Thecvf.com*, pp. 7108–7117, 2021, Accessed: Jul. 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Peng_Sparse-to-Dense_Feature_Matching_Intra_and_Inter_Domain_Cross-Modal_Learning_in_ICCV_2021_paper.html
[9] B. Zhang, S. Zhao, and R. Zhang, "Cross-Domain Semantic Segmentation of Urban Scenes via Multi-Level Feature Alignment," *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 1912–1917, Jan. 2021, doi: 10.1109/icpr48806.2021.9411915.
[10] F. Lv, G. Lin, P. Liu, G. Yang, S. J. Pan, and L. Duan, "Weakly-Supervised Cross-Domain Road Scene Segmentation via Multi-Level Curriculum Adaptation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 9, pp. 3493–3503, Sep. 2021, doi: 10.1109/tcsvt.2020.3040343.
[11] X. Yang, H. Zhou, De Cheng, M. Tian, and N. Wang, "Multi-Level Explicit Feature Alignment Network for Unsupervised Domain Adaptive Person Search," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2026, doi: 10.1109/tcsvt.2026.3704656.
[12] C. Xu, X. Zheng, and X. Lu, "Multi-Level Alignment Network for Cross-Domain Ship Detection," *Remote Sensing*, vol. 14, no. 10, p. 2389, May 2022, doi: 10.3390/rs14102389.
[13] J. Dong, Z. Long, X. Mao, C. Lin, Y. He, and S. Ji, "Multi-Level Alignment Network for Domain Adaptive Cross-Modal Retrieval," *Neurocomputing*, vol. 440, pp. 207–219, Jun. 2021, doi: 10.1016/j.neucom.2021.01.114.
[14] R. Xie, Y. Fei, J. Wang, Y. Wang, and L. Zhang, "Multi-Level Domain Adaptive Learning for Cross-Domain Detection," *Thecvf.com*, pp. 0–0, 2019, Accessed: Jul. 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content_ICCVW_2019/html/TASK-CV/Xie_Multi-Level_Domain_Adaptive_Learning_for_Cross-Domain_Detection_ICCVW_2019_paper.html

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[15] S. Lei, X. Zhang, J. He, F. Chen, B. Du, and C.-T. Lu, "Cross-Domain Few-Shot Semantic Segmentation," *Lecture Notes in Computer Science*, pp. 73–90, 2022, doi: 10.1007/978-3-031-20056-4_5.

[16] Y. Liu, J. Deng, X. Gao, W. Li, and L. Duan, "BAPA-Net: Boundary Adaptation and Prototype Alignment for Cross-Domain Semantic Segmentation," *Thecvf.com*, pp. 8801–8811, 2021, Accessed: Jul. 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Liu_BAPA-Net_Boundary_Adaptation_and_Prototype_Alignment_for_Cross-Domain_Semantic_Segmentation_ICCV_2021_paper.html

[17] J. Luan, J. Ding, and Y. Zhang, "Unsupervised Cross-Domain Radar Target Recognition Using Multilevel Alignment," *IEEE Transactions on Radar Systems*, vol. 3, pp. 630–644, 2025, doi: 10.1109/trs.2025.3560355.

[18] M. Zhang, X. Zhao, W. Li, Y. Zhang, R. Tao, and Q. Du, "Cross-Scene Joint Classification of Multisource Data With Multilevel Domain Adaption Network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 11514–11526, Aug. 2024, doi: 10.1109/tnnls.2023.3262599.

[19] J. Li, F. Ren, M. Liang, J. Yi, and F. Cao, "Cross Domain Object Detection With Multi-Level Domain Feature Refinement," *Computing*, vol. 108, no. 2, Feb. 2026, doi: 10.1007/s00607-026-01621-4.

[20] Y. Li, Z. Li, A. Su, K. Wang, Z. Wang, and Q. Yu, "Semisupervised Cross-Domain Remote Sensing Scene Classification via Category-Level Feature Alignment Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024, doi: 10.1109/tgrs.2024.3392984.

[21] H. Wang, T. Shen, W. Zhang, L.-Y. Duan, and T. Mei, "Classes Matter: A Fine-Grained Adversarial Approach to Cross-Domain Semantic Segmentation," *Lecture Notes in Computer Science*, pp. 642–659, 2020, doi: 10.1007/978-3-030-58568-6_38.

[22] D. Zhao *et al.*, "Learning Pseudo-Relations for Cross-Domain Semantic Segmentation," *Thecvf.com*, pp. 19191–19203, 2023, Accessed: Jul. 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2023/html/Zhao_Learning_Pseudo-Relations_for_Cross-domain_Semantic_Segmentation_ICCV_2023_paper.html

[23] Fengmao Lv, T. Liang, X. Chen, and G. Lin, "Cross-Domain Semantic Segmentation via Domain-Invariant Interactive Relation Transfer," Jun. 2020, doi: 10.1109/cvpr42600.2020.00439.

[24] G. Rui, M. Danelljan, H. Sun, J. D. Mangas, and Van Gool Luc, "Prompting Diffusion Representations for Cross-Domain Semantic Segmentation," arXiv.org. Accessed: Jul. 31, 2026. [Online]. Available: https://arxiv.org/abs/2307.02138

[25] X. Luo, W. Chen, Z. Liang, L. Yang, S. Wang, and C. Li, "Crots: Cross-Domain Teacher–Student Learning for Source-Free Domain Adaptive Semantic Segmentation," *International Journal of Computer Vision*, vol. 132, no. 1, pp. 20–39, Aug. 2023, doi: 10.1007/s11263-023-01863-1.

[26] L. Qing, F. Lv, L. Duan, and B. Gong, "Constructing Self-Motivated Pyramid Curriculums for Cross-Domain Semantic Segmentation: A Non-Adversarial Approach," *Thecvf.com*, pp. 6758–6767, 2019, Accessed: Jul. 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Lian_Constructing_Self-Motivated_Pyramid_Curriculums_for_Cross-Domain_Semantic_Segmentation_A_Non-Adversarial_ICCV_2019_paper.html

[27] Q. Ren, Q. Mao, and S. Lu, "Prototypical Bidirectional Adaptation and Learning for Cross-Domain Semantic Segmentation," *IEEE Transactions on Multimedia*, vol. 26, pp. 501–513, 2024, doi: 10.1109/tmm.2023.3266892.

[28] Q. Zhou *et al.*, "Uncertainty-Aware Consistency Regularization for Cross-Domain Semantic Segmentation," *Computer Vision and Image Understanding*, vol. 221, p. 103448, Aug. 2022, doi: 10.1016/j.cviu.2022.103448.

[29] Y. Gao *et al.*, "PyramidCLIP: Hierarchical Feature Alignment for Vision-Language Model Pretraining," *Advances in Neural Information Processing Systems 35*, pp. 35959–35970, 2022, doi: 10.52202/068431-2606.

[30] H. Hu *et al.*, "Learning Implicit Feature Alignment Function for Semantic Segmentation," arXiv.org. Accessed: Jul. 31, 2026. [Online]. Available: https://arxiv.org/abs/2206.08655

[31] C. Zhao, B. Qin, S. Feng, W. Zhu, L. Zhang, and J. Ren, "An Unsupervised Domain Adaptation Method Towards Multi-Level Features and Decision Boundaries for Cross-Scene Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022, doi: 10.1109/tgrs.2022.3230378.

[32] Y. Liu, J. Wang, C. Huang, Y. Wu, Y. Xu, and X. Cao, "MLFA: Toward Realistic Test Time Adaptive Object Detection by Multi-Level Feature Alignment," *IEEE Transactions on Image Processing*, vol. 33, pp. 5837–5848, 2024, doi: 10.1109/tip.2024.3473532.

[33] J. Xu, Z. Chen, S. Yang, J. Li, H. Wang, and Edith, "MENTOR: Multi-Level Self-Supervised Learning for Multimodal Recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 12, pp. 12908–12917, Apr. 2025, doi: 10.1609/aaai.v39i12.33408.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[34] X. Pei, Y. Gao, X. Li, L. Gao, and X. Zhao, "Multilevel Feature Alignment Method for Advancing Remaining Useful Life Prediction of Rolling Bearing," *IEEE Internet of Things Journal*, vol. 12, no. 12, pp. 22023–22035, Jun. 2025, doi: 10.1109/jiot.2025.3549728.

[35] W. Zhu, C. Zhao, S. Feng, and B. Qin, "Multilevel Feature Alignment Based on Spatial Attention Deformable Convolution for Cross-Scene Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/lgrs.2022.3227251.

[36] J. Nie, Z. Dong, Z. He, H. Wu, and M. Gao, "FAML-RT: Feature Alignment-Based Multi-Level Similarity Metric Learning Network for a Two-Stage Robust Tracker," *Information Sciences*, vol. 632, pp. 529–542, Mar. 2023, doi: 10.1016/j.ins.2023.02.083.

[37] W. Zhou, D. Du, L. Zhang, T. Luo, and Y. Wu, "Multi-Granularity Alignment Domain Adaptation for Object Detection," *Thecvf.com*, pp. 9581–9590, 2022, Accessed: Jul. 31, 2026. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Zhou_Multi-Granularity_Alignment_Domain_Adaptation_for_Object_Detection_CVPR_2022_paper.html

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details