HDRSeg-UDA: Semantic Segmentation for HDR Images with Unsupervised Domain Adaptation

Huei-Yung Lin * and Ming-Yiao Chen

Posted Date: 28 November 2025

doi: 10.20944/preprints202511.2244.v1

Keywords: driving assistance system;deep learning;semantic segmentation;intelligent vehicle;unsupervised domain adaptation;high dynamic range image

Article

HDRSeg-UDA: Semantic Segmentation for HDR Images with Unsupervised Domain Adaptation

Huei-Yung Lin [1,2,†,‡] and Ming-Yiao Chen [1,*,†,‡]

[1] Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 106, Taiwan
[2] Department of Electrical Engineering, National Chung Cheng University, Chiayi 621, Taiwan
* Correspondence: lin@ntut.edu.tw
† Current address: No. 1, Section 3, Zhongxiao E Rd, Da'an District, Taipei City 10608, Taiwan.
‡ These authors contributed equally to this work.

Highlights

What are the main findings?

- The use of HDR images with multi-exposure and feature extraction for road marking semantic segmentation efficiently enables pixel-wise classification on driving images under adverse weather.
- A comprehensive dataset specifically designed for road marking segmentation is introduced, providing a valuable resource for evaluating and improving HDR-based semantic segmentation under different illumination conditions.

What is the implication of the main finding?

- It is feasible to modify the baseline segmentation architecture to better leverage rich features of HDR images with adversarial training and self-training to enhance driving scene understanding tasks.
- The HDR dataset serves as a benchmark for future research in semantic segmentation under various weather conditions.

Abstract

Accurate detection and localization of traffic objects are essential for autonomous driving tasks such as path planning. While semantic segmentation is able to provide pixel-level classification, existing networks often fail under challenging conditions like nighttime or rain. In this paper, we introduce a new training framework that combines unsupervised domain adaptation with high dynamic range imaging. The proposed network uses labeled daytime images along with unlabeled nighttime HDR images. By utilizing the fine details typically lost in conventional SDR images due to dynamic range compression, and incorporating the UDA training strategy, the framework effectively trains a model which is capable of semantic segmentation across the adverse weather conditions. Experiments conducted on four datasets have demonstrated substantial improvements in inference performance under nighttime and rainy scenarios. The accuracy for daytime images is also enhanced through expanded training diversity. Source code is available at https://github.com/ZackChen1140/RMSeg-HDR.

Keywords: driving assistance system; deep learning; semantic segmentation; unsupervised domain adaptation; high dynamic range image; intelligent vehicle

1. Introduction

Semantic segmentation performs pixel-level classification to achieve precise description of object locations and boundaries in an image. By creating dense and structured scene representations, it establishes a core function for advanced perception modules in autonomous systems. When integrated

 doi:10.20944/preprints202511.2244.v1

to advanced driver assistance systems, semantic segmentation is capable of understanding traffic scenes such as lane lines, road markings, and drivable areas. Among these, road marking segmentation is to identify traffic regulations on the road surface, including lane dividers, arrows, and crosswalks, etc., through pixel-wise classification. Precise detection of road markings enhances the stability of ADAS functions, such as lane departure warning, lane keeping, and automatic parking. Furthermore, it provides robust ground image features for accurate localization.

However, achieving semantic segmentation in autonomous driving remains challenging due to diverse real-world factors. First, the variation of traffic regulations in different countries leads to inconsistent road marking patterns, limiting the model generalization across domains. Second, traffic objects such as lane lines and symbols are typically small appeared in images, making accurate pixel-level segmentation difficult. Third, the environmental conditions, including nighttime and rain, often degrade the visibility and increase the recognition uncertainty. On the other hand, recent advances on computation hardware, particularly GPU, TPU, and NPU, enable efficient deployment of deep neural networks on edge devices, facilitating real-time perception and inference.

Semantic segmentation has been developed with deep learning since Long *et al.* introduced fully convolutional networks (FCNs) for end-to-end dense prediction [1]. The subsequent architectures have further enhanced feature representation and contextual reasoning, improving segmentation robustness under complex traffic and weather conditions. Despite the recent progress, most existing investigation on semantic segmentation in traffic scenes primarily targets clear and daytime conditions, while performance under adverse weather still remains limited. The current networks typically rely on supervised learning. It requires large-scale, annotated datasets, which are available for daytime images but insufficient for nighttime or rainy scenes. The scarcity arises from degraded image quality and difficulty of consistent annotation under poor lighting conditions. As a result, model generalization across diverse weather and various illumination remains inadequate.

This paper presents a segmentation framework trained with high dynamic range (HDR) images to enhance robustness under diverse illumination conditions. HDR imaging expands the exposure dynamic range, enabling more accurate representation of both bright and dark regions in the real scene. Characterized by a color depth of 10 bits or higher, HDR images effectively mitigate the distortions from overexposure, underexposure, and low-light noise. To improve cross-domain generalization, unsupervised domain adaptation (UDA) is employed to transfer knowledge from a labeled source domain to an unlabeled target domain. In this study, clear daytime images serve as the source domain, while adverse weather or nighttime images represent the target. By leveraging UDA and HDR, the network achieves consistent semantic segmentation performance across varying illumination and weather conditions, reducing the dependency on extensive manual annotation.

Built upon SegFormer [2], we modify the network architecture for optimizing the feature utilization from HDR images. The proposed technique employs UDA to transfer knowledge from labeled daytime data to unlabeled weather domains, and achieves robust segmentation results across diverse scenarios. Our HDR-based Dual-Path SegFormer is trained and validated using daytime HDR datasets to establish baseline performance. Subsequently, UDA strategies including adversarial learning, self-training, and self-training-based class mixing are incorporated to enhance cross-domain adaptation. They are employed to improve the segmentation robustness under varying weather conditions while preserving class balance and stability during training. The main contributions of this paper are as follows.

- We develop a road marking segmentation model capable of accurate pixel-wise classification on HDR images.
- We demonstrate the feasibility of using HDR images for semantic segmentation under adverse weather.
- The effectiveness of the ClassMix approach for training semantic segmentation model on HDR driving images is verified.
- We establish a new HDR driving dataset for road marking segmentation benchmarking.

 doi:10.20944/preprints202511.2244.v1

2. Related Work

The related work contains three topics: semantic segmentation tasks, unsupervised domain adaptation, and HDR imaging. In discussion of semantic segmentation, it delves into a subtask relevant to autonomous driving: semantic segmentation of road markings. Except for the studies that directly train with HDR images, we also review multi-exposure related techniques.

2.1. Semantic Segmentation

Semantic segmentation is an essential computer vision task. While conventional image classification assigns a single label to an entire image, it classifies every pixel in the image. This allows to describe the object boundary in fine detail, providing both the category of each object in the scene and the associated pixel regions. Most recent semantic segmentation methods are based on deep neural networks. Early approaches mainly rely on CNNs, that enhance feature extraction through convolution operations on images [3,4]. More lately, Transformer-based architectures have become the dominant research trend. Since Dosovitskiy *et al.* [5] demonstrate that Transformer is able to achieve superior classification performance by splitting images into patches, and arranged as sequences of vectors for training, its potential in vision applications has been widely recognized. However, the inherent requirement of significant computation resource makes the Vision Transformers difficult to deploy for more complex tasks. Later variants such as Swin Transformers [6,7] are still not able to achieve real-time segmentation in ADAS systems.

To address these limitations, Xie *et al.* proposed SegFormer [2], which reduces the computational cost of ViT by modifying its backbone network into Mix Transformer and introducing several other improvements. They have demonstrated superior semantic segmentation performance on multiple datasets while maintaining excellent inference speed. Hou *et al.* [8] proposed a knowledge distillation approach called Inter-Region Affinity Knowledge Distillation (IntRA-KD). They decompose the road scene images into different regions labeled as nodes, and then construct an inter-region affinity graph based on similarities in feature distributions among those nodes. This enables a lighter student network to learn the more complex features extracted by the teacher network more effectively. Wu *et al.* introduced a multiscale attention-based dilated CNN [9]. Without increasing the number of parameters, it captures broader semantic information and effectively handles the diverse sizes and shapes of road markings. Hsiao *et al.* [10] proposed a multi-task model incorporating both semantic segmentation and lane detection by leveraging cross-dataset and cross-task learning, using only single or task-limited datasets.

2.2. Unsupervised Domain Adaptation

Most current semantic segmentation studies focus on clear, daytime environments. Although some researchers investigate domain adaptation for more challenging conditions [11–14], it still predominantly relies on SDR images. Nevertheless, we can still adopt and apply to HDR-image-based model training. In the UDA, self-training involves an iterative process where a source-trained Teacher Network generates pseudo-labels for the target domain, which are used to train a Student Network. Updated weights from the Student Network are transferred to the Teacher Network, enabling progressive adaptation. However, inaccurate pseudo-labels can introduce confirmation bias, making verification mechanisms necessary.

Hoyer *et al.* investigated SegFormer-based self-training with several works. DAFormer [11] augments the SegFormer with an Atrous Spatial Pyramid Pooling (ASPP) module [15] and an Exponential Moving Average (EMA) strategy. HRDA [12] extends DAFormer with joint high-/low-resolution processing to capture the fine detail, while MIC [13] further incorporates a Mask Consistency Loss. SePiCo [14] applies Class-Balance Cropping to improve performance on rare categories and show generalizability across networks such as DeepLabv2 [15] and DAFormer. Although they perform well on benchmarks such as Cityscapes→Dark Zurich, many rely on the computationally heavy backbones (e.g., MiT-B5), limiting their practicality for real-time ADAS deployment.

 doi:10.20944/preprints202511.2244.v1

Adversarial training, also known as domain alignment, uses a loss function to encourage a model to learn domain-invariant features. This is typically achieved using a Gradient Reversal Layer (GRL) and a Domain Discriminator. The model attempts to produce similar feature distributions across domains, while the Discriminator tries to identify the domain origin of each feature map, which forms the Min–Max optimization process reminiscent of GANs [16]. Vu *et al.* [17] introduced a Domain Discriminator combined with a minimum-entropy objective to mitigate low-confidence predictions. Wang *et al.* [18] advanced this idea through pixel-level domain discrimination, referred to as Fine-Grained Adversarial Learning. Recently, Cai *et al.* [19] applied SegFormer with unsupervised domain adaptation. By incorporating rare-class sampling, they present a road-marking semantic segmentation model on the multi-weather RLMD-AC dataset [20]. Their model demonstrated strong generalization across diverse weather conditions.

ClassMix is a data augmentation technique which crops and pastes object classes between images to increase data diversity and mitigate class imbalance. Introduced by Olsson *et al.* [21], it has been widely employed in self-training to enhance target-domain variability. Later work, such as DACS [22] and HRDA [12], extended ClassMix into Cross-Domain Mixed Sampling, further enriching data diversity by applying it to both source and target domains. In this research, we employ the core idea of adversarial training and draw inspiration from Fine-Grained Adversarial Learning. The pixel-wise domain discrimination is performed with our modified SegFormer framework.

2.3. Hight Dynamic Range Image

With technology advances, high dynamic range images have become much easier to obtain. By improving the image quality directly from the sensing devices, it is possible to enhance the feature extraction capabilities of deep neural networks. In the ADAS related work, many studies utilize the characteristics of HDR image, specifically the ability to preserve details in both high- and low-brightness regions, for model training. Wang *et al.* [23] developed an HDR-based deep learning model and the training framework capable of detecting vehicle's brake lights more accurately. The HDR-based approach is also adopted for traffic light recognition [24]. Kocdemir *et al.* [25] used HDR images to address the instability issue in conventional images, solving recognition failures which occur in traffic scenes with strong backlighting or overexposed regions.

As the HDR imagery grows the popularity in traffic object detection, researchers begin to explore its potential for traffic-scene semantic segmentation. Weiher employed HDR images captured in virtual environments, converted them to a realistic style, and used to train a semantic segmentation model, which demonstrates the feasibility of using synthetic HDR image for the task [26]. Huang *et al.* [27] used HDR images from the Cityscapes dataset to simulate multiple exposures, generating several SDR images for feature extraction and fusion. It shows strong improvements for several classes in Cityscapes. Beyond using HDR to improve feature extraction, some studies have proposed adjusting image exposure to obtain richer features from a single image. Singh *et al.* [28] introduced an approach that simulates multi-exposure adjustments on a single standard image. This enables the model to extract features across multiple exposure levels, improving object detection performance in dark regions.

Onzon *et al.* proposed a different perspective and approach compared to traditional methods [29]. Conventional methods typically fuse multiple exposure images into a single HDR image using an image signal processor (ISP) before performing object recognition. Since this fusion is usually optimized for human visual perception, it can lead to the loss of critical data for deep neural networks. To address this, they extract features from each exposure image separately and fuse them using the "local cross-attention fusion" method before feeding the result into the detection head for end-to-end object recognition. This method eliminates the need to synthesize a single HDR image and instead directly fuses semantic features from images with different exposures.

 doi:10.20944/preprints202511.2244.v1

3. Method

To deal with semantic segmentation under adverse weather conditions with limited annotated images, this paper presents a UDA-based, end-to-end framework. The proposed network aims to mitigate domain shift between labeled daytime data and unlabeled nighttime and rainy data. More specifically, our architecture incorporates the multi-exposure feature extraction, self-training, and adversarial learning to enhance its robustness and generalization across the illumination changes and weather conditions without the additional annotation on target-domain data. As illustrated in Figure 1, the network model comprises multiple parallel operating modules to optimize domain adaptation and segmentation performance.

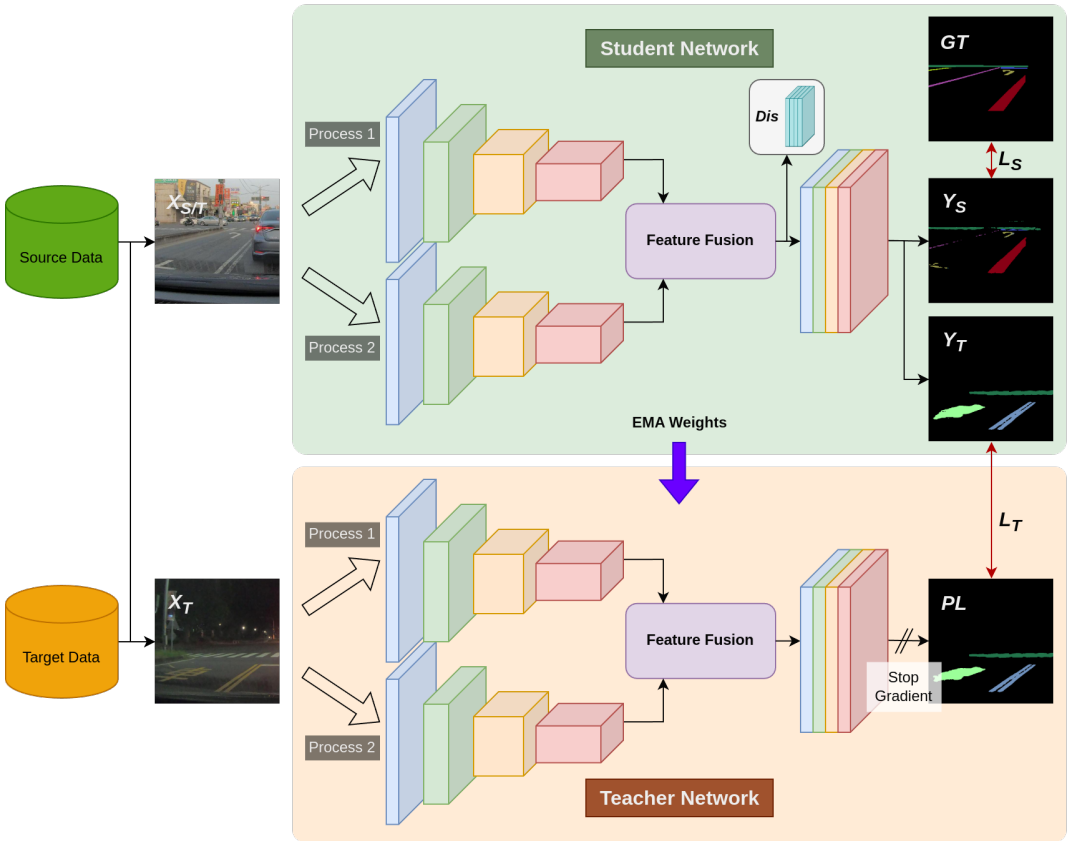


Figure 1. This architecture uses High Dynamic Range (HDR) imaging with multi-exposure feature extraction to reduce information loss in bad weather. It minimizes differences between clear daytime and challenging conditions (night/rainy) by adjusting brightness and employing concurrent self-training and adversarial training in an end-to-end framework.

3.1. Multi-Exposure Feature Extraction for HDR Images

The proposed semantic segmentation network extends SegFormer to improve feature extraction robustness under varying illumination conditions. Multiple SegFormer encoders are employed in parallel to extract the multi-exposure representations from the same input image across different exposure variants. This process produces multiple sets of feature maps which are then integrated into a feature fusion sub-module designed for cross-exposure aggregation and channel compression to reduce the computation while preserving discriminative information. As shown in Figure 2, the input undergoes a series of exposure adjustments, including Gamma correction and log transform, to generate diverse exposure images. They are then processed independently by the corresponding encoders to extract multi-exposure, multi-scale feature maps for downstream fusion and segmentation.

To extract both global and local image characteristics, decoders generate multiple feature maps of varying dimensions. For efficient merge of these maps, especially identically-sized maps from different decoders, both sets are fed into a feature fusion sub-module (see Figure 3). In this sub-module, feature

 doi:10.20944/preprints202511.2244.v1

maps with matching spatial dimensions are first concatenated and then passed through separate multi-layer perceptrons for channel compression. This step enables effective feature blending and alignment of channel dimensions across scales. The smaller feature maps are then upsampled to match the largest one, and all maps are concatenated before being forwarded to the decoder.

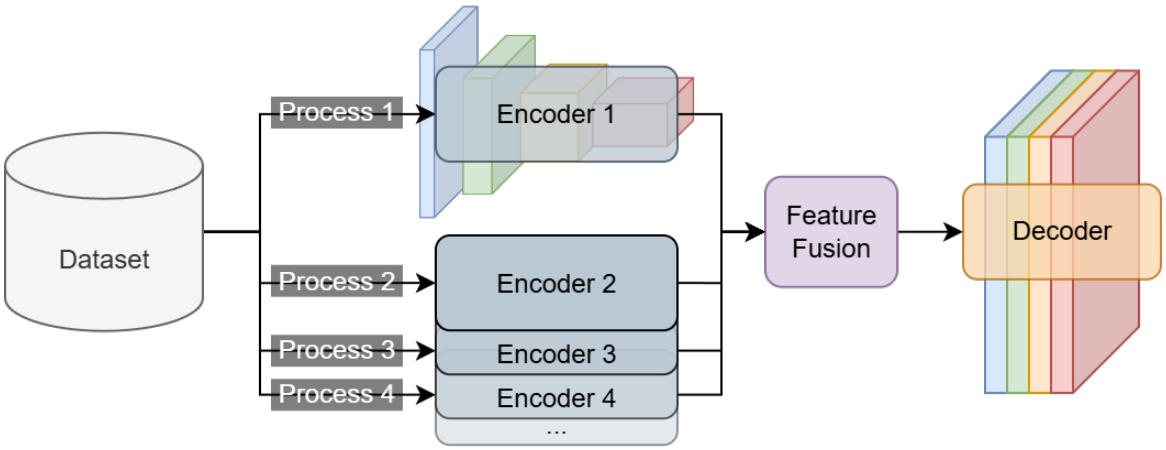


Figure 2. Structure of feature fusion module. This module performs feature fusion on feature maps originating from different decoders. After fusion, the unified features are appropriately scaled and then fed into the decoder.

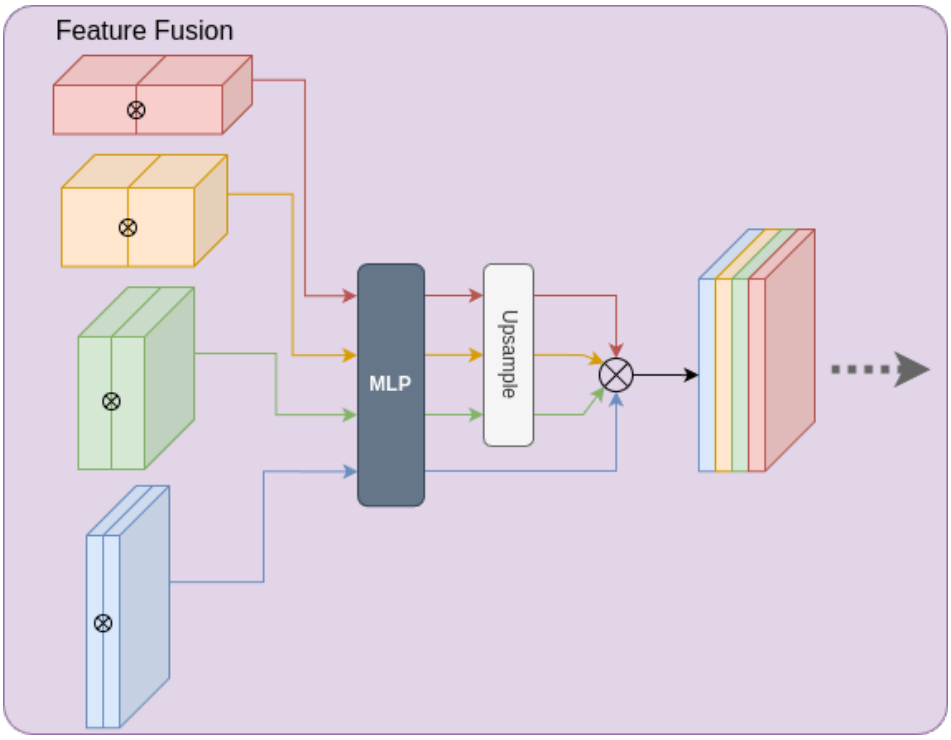


Figure 3. Structure of feature fusion module. This module performs feature fusion on feature maps originating from different decoders. After fusion, the unified features are appropriately scaled and then fed into the decoder.

3.2. Source Domain Supervised Training

This module is trained using source domain data which has been extensively annotated for semantic segmentation. In the source domain, image brightness distributions of conventional images are generally less extreme, and local brightness variations are relatively simple compared to those images in the target domain. Furthermore, the overall image quality remains consistently high. As a result, this module utilizes conventional images for transfer learning, consistent with the common pre-training practices.

 doi:10.20944/preprints202511.2244.v1

The proposed HDRSeg-UDA model is based on SegFormer. To enable the network to extract multi-exposure features in the target domain and enhance its feature extraction capability for the images with varying brightness distributions, we incorporate two SegFormer encoders. These encoders extract features from two images separately with different brightness, and then generate two sets of feature maps to feed into the fusion sub-module for integration. The training of this module utilizes a pixel-wise cross-entropy loss function given by

$$\begin{aligned} L_S^{(i)} &= -\sum_{m=1}^{W}\sum_{n=1}^{H}\sum_{c=1}^{C} y_S^{(i,m,n,c)} \log \tilde{y_S}^{(i,m,n,c)} \\ &= -\sum_{m=1}^{W}\sum_{n=1}^{H}\sum_{c=1}^{C} y_S^{(i,m,n,c)} \log Dec_\phi(Enc_\phi(x_S^{(i)}))^{(m,n,c)} \end{aligned} \tag{1}$$

where $x_S^{(i)}$ represents the i-th image from the source domain, Enc_ϕ and Dec_ϕ denote the encoder and decoder of SegFormer, respectively. ϕ signifies the weights trained using the source domain data. $\tilde{y_S}^{(i)}$ represents the predicted segmentation map inferred by the model, while $y_S^{(i)}$ denotes the ground truth. Finally, H and W represent the height and width of the image, respectively, and C denotes the number of classes.

To enable both decoders to extract distinct image features, the same image undergoes preprocessing to yield two versions with different brightness distributions. This requires a nonlinear adjustment of the image's brightness values, which ensures each emphasizes different luminance characteristics. Thus, we employ exponential and logarithmic functions in this work for nonlinear contrast stretching. The image intensity value subject to exponential contrast stretching accentuates the bright region, while the intensity value processed with a logarithmic function highlights darker feature variations. The computations for the exponential and logarithmic functions are given by

$$\begin{aligned} x_{exp} &= x^\alpha \\ x_{log} &= \frac{\log(1+\beta x)}{\max(\log(1+\beta x))} \end{aligned} \tag{2}$$

where α and β are parameters which control the exponential and logarithmic function curves.

3.3. Target Domain Unsupervised Training

The proposed technique utilizes unlabeled HDR images for training. Since direct image annotations are not available, we employ a separate network model with the identical structure to perform inference on the target domain images and generate pseudo-labels. The pseudo-labels serve as supervision for the original network during its training on the target domain (see Figure 1). Subsequently, the network responsible for producing pseudo-labels in the target domain and the network undergoing domain-specific training are referred to as Teacher and Student Networks, respectively. The generation of pseudo-labels is given by

$$\hat{y}_T^{(j,m,n,c)} = \begin{cases} 1, & c = \arg\max_c Dec_\phi(Enc_\phi(x_T^{(j)}))^{(m,n,c)} \\ 0, & c \neq \arg\max_c Dec_\phi(Enc_\phi(x_T^{(j)}))^{(m,n,c)} \end{cases} \tag{3}$$

where $x_T^{(j)}$ denotes the j-th image from the target domain. Enc_θ and Dec_θ represent the SegFormer encoder and decoder, respectively, where θ indicates the weights of the Teacher Network. $\hat{y}_T$ denotes the pseudo-labels generated by the Teacher Network.

While the Teacher Network's inference capabilities on target domain images are limited, the pseudo-labels cannot substitute ground truth annotations completely. Consequently, it is crucial to apply additional constraints to prevent the confirmation bias. Otherwise, the model would update its weights based on erroneous pseudo-label annotations. In this paper we introduce a confidence

 doi:10.20944/preprints202511.2244.v1

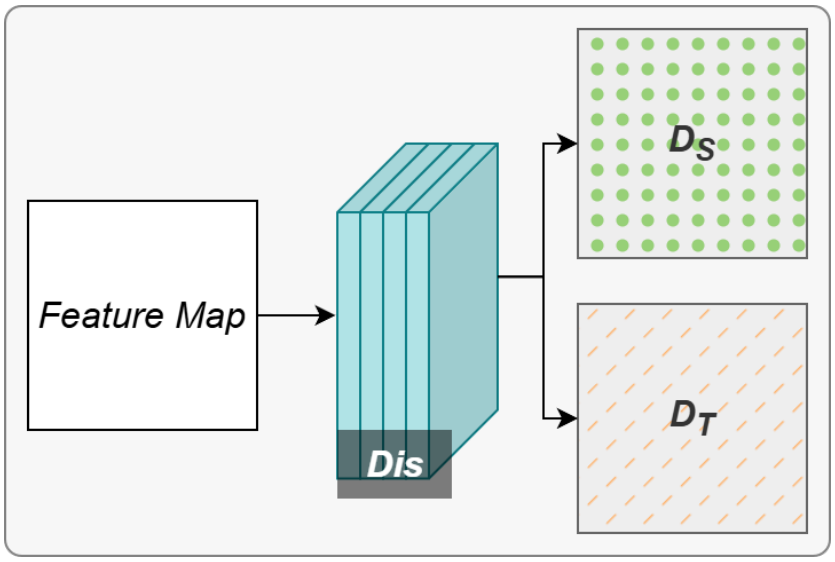


Figure 4. The structure of domain dirscriminator. This architecture incorporates a domain discriminator. Through an adversarial process between the semantic segmentation encoder and the domain discriminator, the encoder's feature extraction capabilities for both source and target domain images are compelled to converge, achieving a more consistent representation across domains.

threshold ϵ when computing the pixel-wise cross-entropy loss between the Student Network's predicted map and the Teacher Network's pseudo-labels. The threshold effectively acts as a mask for pseudo-labels:

$$M^{(j,m,n)} = \begin{cases} 1, & \max_c Dec_\theta(Enc_\theta(x_T^{(j)}))^{(m,n,c)} \geq \epsilon \\ 0, & \max_c Dec_\theta(Enc_\theta(x_T^{(j)}))^{(m,n,c)} < \epsilon \end{cases} \tag{4}$$

If the confidence for all classes at a given pixel in the predicted map does not meet the criterion, the loss is considered.

Finally, the loss function in target domain is calculated by

$$\begin{aligned} L_T^{(j)} &= -\sum_{m=1}^{W}\sum_{n=1}^{H}\sum_{c=1}^{C} y_T^{(j,m,n,c)} \log \tilde{y}_T^{(i,m,n,c)} \\ \log \tilde{y}_T^{(i,m,n,c)} &= M^{(j,m,n)} \log Dec_\theta(Enc_\theta(x_T^{(j)}))^{(m,n,c)} \end{aligned} \tag{5}$$

where $x_T^{(j)}$ represents the j-th image from the target domain, Enc_θ and Dec_θ denote the SegFormer encoder and decoder, respectively. θ indicates the weights of Teacher Network. $\tilde{y}_T^{(j)}$ is the predicted segmentation map inferred by the model, while $y_T^{(j)}$ represents the label derived from the pseudo-label $\hat{y}_T^{(j)}$ and the confidence threshold mask $M^{(j)}$.

To training a model with robust inference capabilities across both domains, it is crucial for sharing the weights ϕ of Student Network to Teacher Network. It ensures Teacher Network can derive more accurate and stable pseudo-labels. In self-training research, there are two common methods to update the weights θ of Teacher Network. The classic approach is to replicate the weights of Student Network directly to Teacher Network, and makes Student as Teacher Network for the next iteration. The other approach is to allow the weights of Teacher Network to gradually converge towards the target domain. In this method, the weights of both Teacher Network and Student Network are combined through a weighted sum to update Teacher Network for the next iteration. We compute using EMA given by

$$\theta_{iter} \leftarrow \alpha\theta_{iter-1} + (1-\alpha)\phi_{iter-1} \tag{6}$$

where α represents the exponentially decay weight, and $iter$ denotes the iteration count.

 doi:10.20944/preprints202511.2244.v1

3.4. Domain Discrimination Training

To facilitate more effective domain shift, beyond the gradual adaptation achieved through self-training, we utilize adversarial learning to fine-tune the model weights. Building upon the supervised learning framework, we develop a pixelwise binary classification decoder by incorporating a domain discriminator, as depicted in Figure 4. The domain discriminator is to infer which domain a given feature map originates from. However, an ideal cross-domain semantic segmentation encoder should produce feature maps that are indistinguishable, regardless of their source domain. Since our domain discriminator is given as a binary classification decoder, the pixelwise binary cross-entropy loss

$$\begin{aligned} L_{Adv}^{(i,j)} &= -\log(1-\tilde{d}_S^{(i)}) - \log \tilde{d}_T^{(j)}) \\ &= -\log(1-Dis(Enc(x_S^{(i)})) - \log(Dis(Enc(x_T^{(j)})) \end{aligned} \tag{7}$$

is employed, where Dis is the domain discriminator, $\tilde{d}_S^{(i)}$ and $\tilde{d}_T^{(j)}$ denote the discrimination maps generated by the domain discriminator after feature extraction from i-th source domain image and j-th target domain image, respectively. In Eq. (8), the source domain is labeled as zero, and the target domain is labeled as one.

Finally, by combining the self-training and adversarial training approaches as illustrated in Figure 1, the loss function for the domain discriminator is written as

$$\begin{aligned} L_{Adv}^{(i,j)} &= -\log(1-\tilde{d}_S^{(i)}) - \log \tilde{d}_T^{(j)}) \\ &= -\log(1-Dis(Enc_\phi(x_S^{(i)})) - \log(Dis(Enc_\theta(x_T^{(j)})) \end{aligned} \tag{8}$$

4. Experiment

4.1. Datasets

We employed two publicly available semantic segmentation datasets, Cityscapes [30] and BDD100K [31], and three public road marking datasets: CeyMo [32], VPGNet [33], and RLMD [33] with the extension RLMD-AC [19]. A major challenge is the lack of HDR images, which are essential for the proposed method, as most datasets (except Cityscapes in clear daytime) provide only 8-bit SDR images. To address this issue, we adopt SingleHDR [34] to reconstruct 32-bit HDR images from single SDR inputs. The resulting images are then stored with 24-bit color depth.

4.1.1. Cityscapes

It is a large-scale urban dataset for semantic segmentation, also providing 16-bit native HDR images (only for clear daytime scenes), The images in this dataset are used as the daytime testing set.

4.1.2. BDD100K

The dataset for large-scale autonomous driving, which covers diverse scenes and weather (daytime, nighttime, rainy). Its subset BDD10K provides 19 semantic classes, manually classified into different weather conditions. All rainy and nighttime images are designated for testing, while daytime images are combined with Cityscapes for training. BDD100K images without annotation are used for augmentation via UDA after being converted to HDR format.

4.1.3. CeyMo

A road marking semantic segmentation dataset collected in Sri Lanka, covering urban, suburban, and rural in daytime, nighttime, and rainy conditions, contains 11 symbolic road marking annotations.

4.1.4. VPGNet

The dataset is for lane line and road marking perception, collected in South Korea. It provides annotations for 18 types of markings with four weather conditions consolidated into three (combining rainy and heavy rain).

 doi:10.20944/preprints202511.2244.v1

4.1.5. RLMD-AC

It is a dataset collected in Taiwan, providing annotated training sets for daytime, and testing sets for various weather conditions, along with unlabeled nighttime and rainy training images. Additional nighttime HDR images are added to the RLMD-AC nighttime training and testing sets to better evaluate the generalization of our method on native HDR data.

4.1.6. Evaluation Metric

In all experiments conducted across the datasets, we employ Mean Intersection over Union (mIoU) as the main evaluation metric. It is important to note that some classes in the BDD10K, CeyMo, and VPGNet datasets are not available in their nighttime or rainy subsets. For these cases, the calculation of mIoU directly excludes these unrepresented classes.

4.2. Implementation

This work modifies the SegFormer architecture as its foundational framework. We employ two encoders of the identical network structure to MiT-B0. This dual-encoder setting allows the model to receive two input images derived from the same source but subjected to different preprocessing. In the feature fusion module (see Figure 3), our implementation modifies the input dimension of the first multi-layer perceptron within the decoder. This modification enables the present of feature maps with twice the original number of channels. Subsequently, the number of channels in these feature maps is compressed back to the original. Finally, the resulting model has approximately 1.93 and 0.52 times the number of parameters in SegFormer-B0 and SegFormer-B1, respectively.

We incorporate ClassMix, as discussed in Section 2.2, into our data augmentation strategy. Furthermore, the approach by Tranheden *et al.* [22], which extends ClassMix beyond mixing exclusively within the target domain is also adopted. However, we observe that using low-confidence pseudo-labels generated during the early training stages with ClassMix often results in incomplete class mixing. To address this issue, our ClassMix strategy is modified to focus on mixing data from the source domain during the initial training phase. As the model training stabilized, we then shift to predominantly mixing data in the target domain.

4.3. Results

The experimental results and comparison with state-of-the-art techniques are shown in Table 1. Compared to the baseline methods (SegFormer [2] and MIC [13]), HDRSeg-UDA shows the best mIoU except the nighttime images in Cityscapes and BDD100K datasets. The consistent improvements on domain changes from clear to night, rainy, and mixed scenarios have validated the effectiveness of our HDR-based UDA framework for multi-weather semantic segmentation. An important finding is the substantial improvement in inference observed under adverse weather, which suppresses the typical degradation of model accuracy in such challenging conditions.

This successful knowledge transfer is attributed to the ability of the UDA framework to adapt semantic understanding from labeled daytime (source) images to unlabeled adverse-weather (target) domains. Moreover, the accuracy is also increased on daytime datasets. This demonstrates that our proposed multi-exposure feature extraction module (i.e. Dual-Path SegFormer) designed to exploit HDR inputs yields more robust and stable features even under normal illumination. The experiments on multiple datasets with adverse weather, including Cityscapes, BDD100K, CeyMo, VPGNet, and RLMD-AC, show our integrated training approach leveraging self-training and adversarial training mechanisms is a highly generalizable semantic segmentation model.

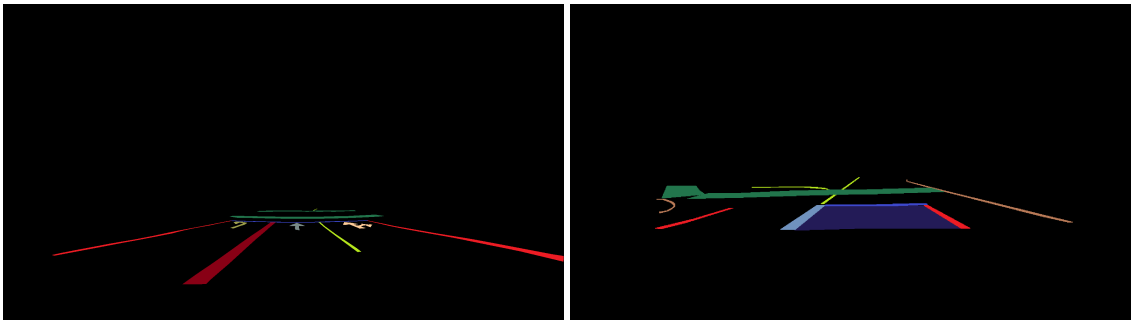
 doi:10.20944/preprints202511.2244.v1

Table 1. The comparison of our HDRSeg-UDA with state-of-the-art methods. The performance is reported as mIoU in %.

Task	Method	Cityscapes & BDD100K			CeyMo		
		clear ↑	night ↑	rainy ↑	clear ↑	night ↑	rainy ↑
Clear → Night	Baseline	-	-	-	71.33	36.29	-
	HDRSeg-UDA	-	-	-	**76.99**	**75.44**	-
Clear → Rainy	Baseline	-	-	-	71.33	-	61.70
	HDRSeg-UDA	-	-	-	**76.42**	-	**80.56**
Clear → Mixed	Baseline	57.83	21.47	34.74	71.33	36.29	61.70
	MIC [13]	61.22	**29.18**	38.73	68.28	59.12	73.68
	HDRSeg-UDA	**65.55**	28.33	**40.13**	**77.03**	**74.55**	**79.21**

Task	Method	VPGNet			RLMD-AC		
		clear ↑	night ↑	rainy ↑	clear ↑	night ↑	rainy ↑
Clear → Night	Baseline	33.98	29.29	-	52.36	28.53	-
	HDRSeg-UDA	**34.01**	**32.91**	-	**55.99**	**37.72**	-
Clear → Rainy	Baseline	33.98	-	29.59	52.36	-	32.32
	HDRSeg-UDA	**34.35**	-	**36.82**	**53.83**	-	**37.79**
Clear → Mixed	Baseline	33.98	29.29	29.59	52.36	28.53	32.32
	MIC [13]	32.79	23.15	30.75	53.03	35.43	39.60
	HDRSeg-UDA	**35.80**	**35.58**	**37.28**	**57.88**	**40.06**	**40.93**

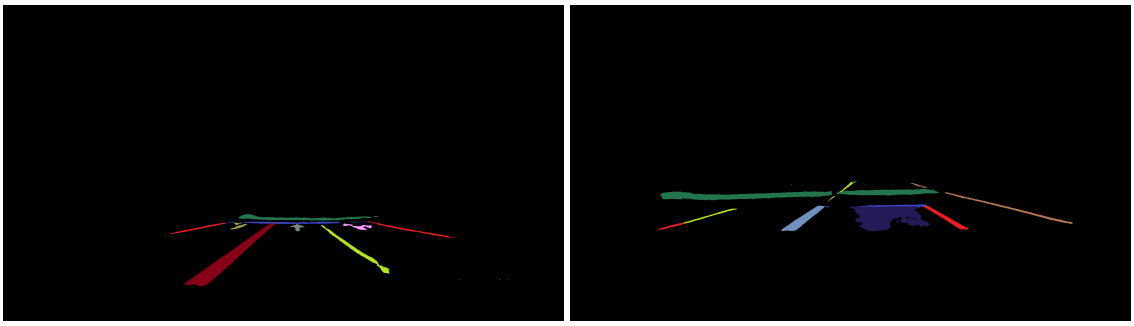
(**a**) Image



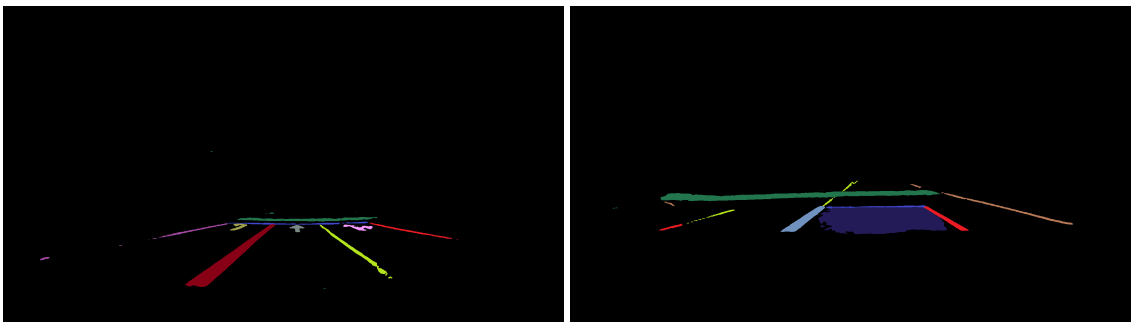
(**b**) Ground Truth

Figure 5. *Cont.*

 doi:10.20944/preprints202511.2244.v1



(**c**) Baseline



(**d**) HDRSeg-UDA

Figure 5. The results for the target domain on the RLMD-AC dataset.

4.4. Ablation Study

To assess the contribution of each component of HDRSeg-UDA, the ablation experiments are conducted on the RLMD-AC dataset to evaluate the Dual-Encoder, Triple-Encoder, Self-Training, Discriminator, and ClassMix modules, as tabulated in Table 2. We further examine the impact of HDR imagery by comparing model performance when trained and tested with SDR versus HDR images. As the evaluation tabulated in Table 3 for Cityscapes and RLMD-AC datasets, the effectiveness of each key component, including the UDA techniques (self-training and adversarial training) and the use of high-bit-depth HDR data, highlights the overall robustness of the integrated approach across multiple image datasets.

Table 2. Contribution of each submethod in the RLMD-AC CLEAR→NIGHT task. The performance is reported as mIoU in %.

Method	Clear ↑	Night ↑	Rainy ↑
Baseline	52.62	25.68	32.86
Base + Dual-Encoder	53.47	25.58	33.63
Base + Triple-Encoder	54.12	29.06	33.53
Base + DE + Self-Training	54.67	26.27	36.52
Base + DE + ST + Discriminator	54.97	31.66	38.04
Base + DE + ST + Dis + ClassMix	57.88	40.06	40.93

 doi:10.20944/preprints202511.2244.v1

Table 3. Contribution of different bit depths in our architecture.

Datasets	Weather	SDR (8-bit) ↑	HDR (16/24-bit) ↑
Cityscapes	Clear	62.09	65.55
RLMD-AC	Night	29.30	32.26

5. Conclusions

This paper presents a new training framework that enhances semantic segmentation for autonomous driving, particularly in challenging nighttime and rainy conditions. By incorporating HDR imagery with UDA techniques (self-training and adversarial training), our proposed HDRSeg-UDA leverages labeled daytime images and unlabeled nighttime HDR data effectively. The integration of the HDR feature representation with UDA generalization capabilities results in a robust model exhibiting excellent road marking segmentation under different scanarios. Experiments conducted on four adverse weather datasets have demonstrated the substantial improvements in inference accuracy on both nighttime and rainy weather conditions, and also yielding a notable boost in accuracy on daytime scenes.

Author Contributions: Conceptualization, Huei-Yung Lin and; methodology, Huei-Yung Lin and Ming-Yiao Chen; software, Ming-Yiao Chen; validation, Ming-Yiao Chen; formal analysis, Ming-Yiao Chen; investigation, Huei-Yung Lin; resources, Huei-Yung Lin; data curation, Ming-Yiao Chen; writing—original draft preparation, Huei-Yung Lin; writing—review and editing, Huei-Yung Lin; visualization, Ming-Yiao Chen; supervision, Huei-Yung Lin; project administration, Huei-Yung Lin; funding acquisition, Huei-Yung Lin; All authors have read and agreed to the published version of the manuscript.

Funding: The support of this work is in part by the National Science and Technology Council of Taiwan under Grant 109-2221-E-194-037-MY3.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The code and dataset presented in the study are openly available at https://github.com/ZackChen1140/RMSeg-HDR.

Acknowledgments: The support of this work in part by the National Science and Technol- ogy Council of Taiwan under Grant 109-2221-E-194-037-MY3 is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TLD	Traffic Light Detection
C2I	Car-to-Infrastructure
CNN	Convolutional Neural Network
SOTA	State-Of-The-Art
SimAM	Simple Attention Module
ECA	Efficient Channel Attention Mechanism
CIoU	Complete Intersection of Union
EIoU	Efficient Intersection of Union
HSM	Hard Sample Mining
GFLOPs	Giga Floating-point Operations Per Second

 doi:10.20944/preprints202511.2244.v1

References

1. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
2. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems* **2021**, *34*, 12077–12090.
3. Gidaris, S.; Komodakis, N. Object detection via a multi-region and semantic segmentation-aware cnn model. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2015, pp. 1134–1142.
4. Tokunaga, H.; Teramoto, Y.; Yoshizawa, A.; Bise, R. Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 12597–12606.
5. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* **2020**.
6. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10012–10022.
7. Wang, J.; Liao, X.; Wang, Y.; Zeng, X.; Ren, X.; Yue, H.; Qu, W. M-SKSNet: Multi-scale spatial kernel selection for image segmentation of damaged road markings. *Remote Sensing* **2024**, *16*, 1476.
8. Hou, Y.; Ma, Z.; Liu, C.; Hui, T.W.; Loy, C.C. Inter-region affinity distillation for road marking segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12486–12495.
9. Wu, J.; Liu, W.; Maruyama, Y. Automated road-marking segmentation via a multiscale attention-based dilated convolutional neural network using the road marking dataset. *Remote Sensing* **2022**, *14*, 4508.
10. Hsiao, H.C.; Cai, Y.C.; Lin, H.Y.; Chiu, W.C.; Chan, C.T.; Wang, C.C. FuseRoad: Enhancing Lane Shape Prediction Through Semantic Knowledge Integration and Cross-Dataset Training. In Proceedings of the 2025 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2025, pp. 897–902.
11. Hoyer, L.; Dai, D.; Van Gool, L. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 9924–9935.
12. Hoyer, L.; Dai, D.; Van Gool, L. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation. In Proceedings of the European conference on computer vision. Springer, 2022, pp. 372–391.
13. Hoyer, L.; Dai, D.; Wang, H.; Van Gool, L. MIC: Masked image consistency for context-enhanced domain adaptation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 11721–11732.
14. Xie, B.; Li, S.; Li, M.; Liu, C.H.; Huang, G.; Wang, G. Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2023**, *45*, 9004–9021.
15. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *40*, 834–848.
16. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Communications of the ACM* **2020**, *63*, 139–144.
17. Vu, T.H.; Jain, H.; Bucher, M.; Cord, M.; Pérez, P. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 2517–2526.
18. Wang, H.; Shen, T.; Zhang, W.; Duan, L.Y.; Mei, T. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 642–659.
19. Cai, Y.C.; Hsiao, H.C.; Chiu, W.C.; Lin, H.Y.; Chan, C.T. RMSeg-UDA: Unsupervised Domain Adaptation for Road Marking Segmentation Under Adverse Conditions. In Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2025, pp. 13471–13477.

 doi:10.20944/preprints202511.2244.v1

20. Hsiao, H.C.; Cai, Y.C.; Lin, H.Y.; Chiu, W.C.; Chan, C.T. RLMD: A Dataset for Road Marking Segmentation. In Proceedings of the 2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan). IEEE, 2023, pp. 427–428.
21. Olsson, V.; Tranheden, W.; Pinto, J.; Svensson, L. Classmix: Segmentation-based data augmentation for semi-supervised learning. In Proceedings of the Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021, pp. 1369–1378.
22. Tranheden, W.; Olsson, V.; Pinto, J.; Svensson, L. Dacs: Domain adaptation via cross-domain mixed sampling. In Proceedings of the Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021, pp. 1379–1389.
23. Wang, J.G.; Zhou, L.; Song, Z.; Yuan, M. Real-time vehicle signal lights recognition with HDR camera. In Proceedings of the 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2016, pp. 355–358.
24. Wang, J.G.; Zhou, L.B. Traffic light recognition with high dynamic range imaging and deep learning. *IEEE Transactions on Intelligent Transportation Systems* **2018**, *20*, 1341–1352.
25. Kocdemir, I.H.; Akyuz, A.O.; Koz, A.; Chalmers, A.; Alatan, A.; Kalkan, S. Object detection for autonomous driving: high-dynamic range vs. low-dynamic range images. In Proceedings of the 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 2022, pp. 1–5.
26. Weiher, M. Domain adaptation of HDR training data for semantic road scene segmentation by deep learning **2019**.
27. Huang, T.; Song, S.; Liu, Q.; He, W.; Zhu, Q.; Hu, H. A novel multi-exposure fusion approach for enhancing visual semantic segmentation of autonomous driving. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* **2023**, *237*, 1652–1667.
28. Singh, K.; Parihar, A.S. MRN-LOD: multi-exposure refinement network for low-light object detection. *Journal of Visual Communication and Image Representation* **2024**, *99*, 104079.
29. Onzon, E.; Bömer, M.; Mannan, F.; Heide, F. Neural exposure fusion for high-dynamic range object detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 17564–17573.
30. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.
31. Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; Darrell, T. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2636–2645.
32. Jayasinghe, O.; Hemachandra, S.; Anhettigama, D.; Kariyawasam, S.; Rodrigo, R.; Jayasekara, P. Ceymo: See more on roads-a novel benchmark dataset for road marking detection. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3104–3113.
33. Lee, S.; Kim, J.; Shin Yoon, J.; Shin, S.; Bailo, O.; Kim, N.; Lee, T.H.; Seok Hong, H.; Han, S.H.; So Kweon, I. Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017, pp. 1947–1955.
34. Liu, Y.L.; Lai, W.S.; Chen, Y.S.; Kao, Y.L.; Yang, M.H.; Chuang, Y.Y.; Huang, J.B. Single-image HDR reconstruction by learning to reverse the camera pipeline. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 1651–1660.

 doi:10.20944/preprints202511.2244.v1

Short Biography of Authors

Huei-Yung Lin is a professor at the Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei, Taiwan. He also has a joint appointment with the Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan. Professor Lin received his Ph.D. degree in electrical and computer engineering from State University of New York at Stony Brook, US. In 2002 he joined National Chung Cheng University, Taiwan, as an assistant professor and was promoted to a full professor in 2013. He served as the Director of Research Liaison Division from 2009 to 2013 and the Director of Academic Development Division from 2012 to 2014, both positions are with the Office of Research and Development, National Chung Cheng University. He is an author of over 230 international conference and journal papers and 9 book chapters. Professor Lin holds 11 patents for invention in the United States and 9 patents in Taiwan. He also serves as an organizer committee member and a program committee member of more than 60 international conferences. Dr. Lin is the recipient of the Excellent Research Award from National Chung Cheng University, the Outstanding Academic-Industry Cooperation Award from the Taiwan Association of System and Science and Engineering (TASSE) and the Outstanding Robotics Engineer Award from Robotics Society of Taiwan (RST). His research interests include machine learning, computer vision, robotics, and mechatronics. He is a fellow of IET and RST, and a senior member of IEEE and Optica.

Ming-Yiao Chen received the B.S. degree from Yuan Ze University, Taoyuan, Taiwan and the M.S. degree from National Taipei University of Technology, Taipei, Taiwan, both in in computer science and information engineering. His research interests include embedded system design, machine learning, image processing, intelligent vehicles, robotics. He is now with Gemtek Ltd., Taiwan, as a senior engineer working on algorithm development and embedded system design.