Real-time Domain Adaptation in Semantic Segmentation

Ianniello Luca, Martone Raffaele, Sirica Antonio

January 27, 2025

Abstract

This study addresses the problem of real-time domain adaptation in semantic segmentation, focusing on bridging performance gaps between source and target domains. Leveraging the LoveDA dataset, which encompasses rural and urban domains, we evaluate the performance of state-of-the-art models, including DeepLabv2 and PIDNet, under domain adaptation settings. We explore a range of adaptation techniques, such as data augmentation, adversarial learning, domain adaptation via cross-domain mixed sampling (DACS), Prototype-based Efficient MaskFormer (PEM), and unpaired image-to-image translation with CycleGAN. Our findings reveal the strengths and limitations of these methods in improving segmentation performance across domains while preserving computational efficiency.

1 Introduction

Semantic segmentation is a fundamental task in computer vision that involves partitioning an image into semantically meaningful regions, typically corresponding to different object classes. This task is crucial for various applications, including autonomous driving, medical imaging, and remote sensing. Recent advancements in deep learning have significantly improved the performance of semantic segmentation models [4].

While deep learning models such as DeepLab [2] and PIDNet [3] have achieved remarkable performance on benchmark datasets, their generalization to unseen domains remains a significant obstacle. Domain shifts, arising from differences in image characteristics, environments, and data distributions, often result in substantial performance degradation.

Domain adaptation aims to bridge the performance gap between source and target domains, enabling models trained on one domain to generalize effectively to another. The LoveDA dataset [7], with its distinct rural and urban domain settings, provides a robust benchmark for evaluating domain adaptation techniques in semantic segmentation.

This project investigates real-time domain adaptation techniques for semantic segmentation, focusing on the LoveDA dataset. We evaluate state-of-the-art models, such as DeepLabv2 and PIDNet, under various domain adaptation scenarios. Specifically, we analyze the challenges of adapting models from rural to urban domains and vice versa, and we explore solutions including data augmentation, adversarial training [6], image-to-image translation approaches (DACS) [5], real-time networks (PEM) [1], and style transfer preprocessing models, like CycleGAN [8]. By experimenting with these approaches, we aim to identify effective methods for improving domain adaptation performance while maintaining real-time capabilities.

2 Related Work

In this section, we provide an overview of the state-of-the-art models and techniques for semantic segmentation and domain adaptation used to implement the basic structure of our project.

2.1 DeepLabv2

DeepLab has played a pivotal role in advancing semantic segmentation by overcoming key challenges in traditional deep convolutional neural networks (DCNNs), including resolution loss, scale variability, and poor boundary localization [2]. DeepLab V2 specifically addresses these issues through three major contributions: atrous convolution, Atrous Spatial Pyramid Pooling (ASPP), and Fully Connected Conditional Random Fields (CRFs). Atrous convolution, or dilated convolution, mitigates resolution loss by expanding the receptive field without increasing computational costs or reducing spatial resolution. This enables accurate pixel-level predictions while preserving fine details. To tackle scale variability, the ASPP module aggregates multi-scale contextual information using parallel atrous convolutions with varying dilation rates, ensuring robustness to objects appearing at different sizes. Finally, Fully Connected CRFs refine segmentation results by modeling pixel relationships based on

spatial proximity and color similarity. This post-processing step enhances boundary delineation and recovers intricate object edges. By addressing these core challenges, DeepLab V2 has become a benchmark framework in semantic segmentation, achieving high-resolution predictions, robust multi-scale representations, and precise boundary localization. Its innovations have significantly influenced subsequent research, setting a new standard for segmentation methodologies. DeepLabv2 was used in our project in the first experiments to evaluate its performance in domain adaptation scenarios and to better understand the behaviour of a classic segmentation network.

2.2 PIDNet

PIDNet is a real-time semantic segmentation network inspired by Proportional-Integral-Derivative (PID) controllers, which are commonly used in control systems to achieve precision and stability [3]. By drawing from PID control theory, PIDNet introduces a unique architecture with three specialized branches—P (proportional), I (integral), and D (derivative)—each capturing distinct and complementary information. The P-branch focuses on preserving spatial details, the I-branch aggregates global contextual information, and the D-branch enhances boundary precision. This design allows PIDNet to achieve a balance between low-latency processing and high segmentation quality. Particularly effective in real-time applications such as autonomous driving and robotics, PIDNet delivers state-of-the-art performance while maintaining a lightweight structure. This efficiency makes it suitable for resource-constrained environments without significant compromise in accuracy, showcasing its versatility and practical utility across diverse domains. PIDNet is the basic model we will use for our experiments and we have also implemented some changes to improve its performance in domain adaptation scenarios. These changes will be discussed in the following sections.

2.3 LoveDa Dataset

The dataset used in our experiments is the LoveDA dataset. The LoveDA dataset is a novel benchmark for domain adaptation in semantic segmentation, featuring distinct rural and urban domains with varying environmental characteristics [7]. This dataset provides a challenging testbed for evaluating the generalization capabilities of segmentation models across diverse settings. By encompassing rural and urban scenes, LoveDA captures the complexity of real-world applications, where domain shifts are prevalent and pose significant challenges to model performance. The dataset's diverse landscapes, lighting conditions, and object distributions necessitate robust adaptation techniques to ensure accurate segmentation results.

3 Methodology

In this section we describe the methodology used to implement the project. We will discuss the models used, the techniques applied, and the implemented variation conducted to evaluate the performance of the models in domain adaptation scenarios.

3.1 DeepLab2 implementation

3.2 PIDNet implementation

The PIDNet model used in this project was originally proposed by [3] and was adapted by us for use with the LoveDA dataset. For our implementation, we selected the PIDNet small model, utilizing a pretrained ImageNet model for initialization. While the core structure of PIDNet remained unchanged, we introduced several modifications to enhance its performance in domain adaptation scenarios. Specifically, we experimented with various loss functions, optimizers, and the integration of a learning rate scheduler. These adjustments were aimed at improving the model's ability to generalize across domains and better handle the inherent challenges of domain shifts in the dataset.

The different losses we have implemented are the following:

- Cross Entropy: This is the standard loss function for semantic segmentation tasks. It calculates the pixel-wise difference between the predicted and true class probabilities, penalizing incorrect classifications. While effective, it may struggle with class imbalance in datasets. This loss function was just defined in the PIDNet model.
- OHEM Cross Entropy : This variant of the cross entropy loss prioritizes hard-to-classify pixels during training. By focusing on these challenging examples, Ohem enhances the model's ability to learn from difficult cases, improving overall segmentation quality. Also this loss function was just defined in the PIDNet model.
- Boundary Loss: This loss function focuses on the boundary regions of objects, enhancing the model's ability to capture fine details and object edges. By emphasizing boundary pixels, the model can achieve more precise segmentation results.

- Focal Loss: Focal Loss further addresses class imbalance by reducing the impact of well-classified pixels and focusing on hard-to-classify ones. This dynamic adjustment of loss contributions makes it especially effective in datasets with highly imbalanced classes.

- Dice Loss: This type of loss was also used in the DeepLabV2 implementation. Designed to handle class imbalance, Dice Loss measures the overlap between predicted and ground truth masks. By optimizing the Dice coefficient, this loss improves performance on underrepresented classes and ensures a balanced segmentation output.

The different optimizers we have implemented are the following:

- Stochastic Gradient Descent (SGD): This classic optimizer updates model parameters based on the gradient of the loss function. While effective, SGD may struggle with noisy gradients and slow convergence. To address these limitations, PIDNet leverages SGD with momentum, which accumulates gradients over time to stabilize updates and accelerate learning. This scheduler was used in the standard PIDNet model implementation thanks the optimizer library of PyTorch.

- ADAM: This adaptive optimizer dynamically adjusts the learning rate for each parameter based on moment estimates. It is particularly effective in scenarios where the learning rate needs to adapt quickly.

In addition, we have defined a Cosine Annealing Learning Rate Scheduler to dynamically adjust the learning rate throughout training. This scheduler gradually reduces the learning rate following a cosine curve, from an initial maximum value to a defined minimum value. To enhance the training process, a warm-up phase is incorporated at the start, where the learning rate linearly increases to the base learning rate over a specified number of epochs. In our case, we choose five epochs considering that all the training done was on 20 epochs. Beyond the warm-up phase, the cosine annealing strategy begins, promoting smooth convergence by allowing the learning rate to decrease progressively. This approach ensures that the model benefits from rapid initial learning, followed by gradual fine-tuning as training progresses, avoiding abrupt changes that could destabilize optimization. This combination of flexibility and stability contributes to improved generalization and segmentation performance.

In the experiments and result section we will discuss the performance of the model with the different loss functions, with different optimizer and the impact of the learning rate scheduler on the performance of the model.

3.3 Data Augmentation

Data augmentation is a crucial technique in semantic segmentation, particularly when dealing with domain adaptation. It involves generating additional training data by applying various transformations to the existing dataset. In our specific project, we have trained PIDNet in the source urban domain and evaluated it in the target rural domain. The results are shown in the Experimental and Result section (4). To improve performance and apply Domain Adaptation, we have implemented three different data augmentation techniques:

- The first type of data augmentation is composed by a Random Brightness and Contrast transformation and a Random Shadow. The first technique randomly adjusts the brightness and contrast of images, simulating variations in lighting conditions. The second technique introduces random shadows to images, enhancing the model's ability to detect objects under different illumination settings.

- The second type of data augmentation is composed by a Hue Saturation Value (HSV) transformation and a Gaussian Blur. The first technique randomly alters the hue, saturation, and value of images, creating diverse color variations. The second technique applies Gaussian blur to images, smoothing out pixel noise and enhancing object contours.

- The third type of data augmentation is focused on geometrical augmentation. This technique includes a Horizontal Flip, a Vertical Flip and a Random Rotation of 90 degrees. The first two techniques flip images horizontally and vertically, respectively, to introduce spatial variations. The third technique rotates images randomly by 90 degrees, simulating different orientations of objects in the scene.

All these techniques are followed by a normalization of the images. The results of the experiments with the data augmentation techniques are shown in the Experimental and Result section (4).

3.4 Adversarial Learning

3.5 Domain Adaptation via Cross-domain Mixed Sampling (DACS)

Another way to improve the performance of the model in domain adaptation scenarios is to use the

Domain Adaptation via Cross-domain Mixed Sampling (DACS) technique. This approach encourages the model to learn domain-invariant features that are robust to domain shifts, thereby reducing the domain gap between the source and target domains. By combining source and target domain samples, DACS enables seamless adaptation to new environments.

In our implementation, the urban domain serves as the source domain, while the rural domain is treated as the target domain. We applied DACS by mixing source images and labels with target images and pseudo-labels during training. The pseudo-labels for the target domain are derived by applying the argmax operation on the logits generated by the basic model. This mixing strategy ensures the model learns transferable features across diverse environments.

To compute the training loss, we combined the source loss and the mixed loss, weighted by a factor. This iterative approach ensures the model learns robust, domain-invariant features. Detailed experimental results, along with an evaluation of the impact of DACS, are presented in the Experimental and Results section (4).

3.6 Prototype-based Efficient MaskFormer (PEM)

3.7 Unpaired Image-to-Image Translation with CycleGAN

4 Experiments and results

In this section, we present the experimental results obtained by evaluating the performance of the models in domain adaptation scenarios. We analyze the impact of different techniques, such as data augmentation, adversarial learning, DACS, PEM, and CycleGAN, on segmentation quality and domain adaptation capabilities. All the tests are done considering 20 epochs for the training phase. The fixed parameters for all the experiments are the learning rate for ADAM optimizer, that is 0.001, the learning rate for SGD optimizer, that is 0.01, and the batch size for GPUs that is equal to 6. The results are discussed in terms of mean Intersection over Union, latency, FLOPs and number of parameters.

4.1 DeepLabv2 Experiments

4.2 PIDNet Experiments

As described in the Methodology section, we have implemented different loss functions, optimizers and a learning rate scheduler to improve the performance of the PIDNet model in domain adaptation scenarios. In this subsection we report all the results obtained with different combination of The results of the experiments are shown in the following tables.

4.3 Data Augmentation Experiments

Considering the previous better performing model, we have trained the found PIDNet con

As described in the Methodology section, we have implemented three different data augmentation techniques to improve the performance of the PIDNet model in domain adaptation scenarios. In this subsection we report all the results obtained with the different data augmentation techniques. The results of the experiments are shown in the following tables.

4.4 Adversarial Learning Experiments

4.5 DACS Experiments

The result obtained with the DACS technique, done on the optimal configuration

4.6 PEM Experiments

4.7 CycleGAN Experiments

5 Conclusion

References

[1] Niccolo Cavagnero, Gabriele Rosi, Claudia Cuttano, Francesca Pistilli, Marco Ciccone, Giuseppe Averta, and Fabio Cermelli. Pem: Prototype-based efficient maskformer for image segmentation. *arXiv preprint arXiv:2402.19422v3*, 2024. 1

[2] Liang-Chieh Chen, George Papandreou, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 1

[3] Hao Feng et al. Pidnet: A real-time semantic segmentation network inspired by pid controllers. *arXiv preprint arXiv:2103.12370*, 2021. 1, 2

[4] Shijie Hao, Yuan Zhou, and Yanrong Guo. A brief survey on semantic segmentation with deep learning. *arXiv preprint arXiv:2007.04063*, 2020. 1

[5] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021. 1

[6] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[7] Yucheng Wang et al. Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *arXiv preprint arXiv:2110.08733*, 2021. 1, 2

[8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593v7*, 2020. 1