Real-time Domain Adaptation in Semantic Segmentation

TA: Claudia Cuttano (claudia.cuttano@polito.it)
TA: Gabriele Rosi (gabriele.rosi@polito.it)

STARTING REPOSITORY

https://github.com/Gabrysse/MLDL2024_project1

OVERVIEW

The main objective of this project is to become familiar with the task of Domain Adaptation applied to the Real-time Semantic Segmentation networks. The student should understand the general approaches to perform Domain Adaptation in Semantic Segmentation and the main reason to apply them to real-time networks. Before starting, the student should read [1] [2] [3] to get familiar with the tasks. As the next step, the student should train the real-time segmentation network [2] on the target dataset [4] to define the upper bound. Then, he/she should train the network [2] on the synthetic dataset [5] and evaluate the drop of the performance when directly testing the trained model on the target images [4]. For the last part of the project, the student should implement a variation for the project.

1st STEP) RELATED WORKS

Reading paper to get familiar with the task

Before starting it is mandatory to take time to familiarize yourself with the tasks of Semantic Segmentation, Domain Adaptation and Real-time Semantic Segmentation. It is compulsory to understand what are the main problems and the main solutions to tackle them in literature. More in detail, read:

- [1] to understand Semantic Segmentation;
- [2][3] to understand classic and real-time solutions for Semantic Segmentation;
- [4] to get familiar with the several solutions to perform unsupervised domain adaptation in Semantic Segmentation;
- [5] [6] to get familiar with the datasets that will be used in this project;

2nd STEP) TESTING SEMANTIC SEGMENTATION NETWORKS

a) **Classic semantic segmentation network.**

For this step, you have to train a classic segmentation network (DeepLabV2 [2]) on the Cityscapes dataset.

- *Dataset*: Cityscapes [5]
- *Training epochs*: 50
- *Training resolution (Cityscapes)*: 1024x512
- *Test resolution (Cityscapes)*: 1024x512
- *Backbone*: R101 (pre-trained on ImageNet) [2]
- *Semantic classes*: 19
- *Metrics*: Mean Intersection over Union (mIoU) [read this to understand the metrics], latency, FLOPs, number of parameters.

Table 1) Classic Cityscapes	**mIoU (%)**	**Latency**	**FLOPs**	**Params**
DeepLabV2 - 50 epochs				

b) Real-time semantic segmentation network.

For this step, you have to train a real-time segmentation network (BiSeNet [3]) on the Cityscapes dataset.

- *Dataset*: Cityscapes [5]
- *Training epochs*: 50
- *Training resolution (Cityscapes)*: 1024x512
- *Test resolution (Cityscapes)*: 1024x512
- *Backbone*: ResNet18 (pre-trained on ImageNet) [3]
- *Semantic classes*: 19
- *Metrics*: mIoU, latency, FLOPs, number of parameters.

Table 2) Real-time Cityscapes	**mIoU (%)**	**Latency**	**FLOPs**	**Params**
BiSeNet - 50 epochs				

3rd STEP) DOMAIN SHIFT

From now on, we will employ BiSeNet as our segmentation to ease the resource requirements of the next experiments.

Consider as upper bound the results obtained in Table 2, i.e. the segmentation networks trained on the labeled target images (Cityscapes).

a) Evaluating the domain shift problem in Semantic Segmentation

In semantic segmentation collecting manually annotated images is expensive. A popular solution consists in adopting synthetic datasets (i.e. artificial images generated in a simulation environment).

Specifically, in this step we employ the synthetic images from GTA5 [5] (source domain) to train our real-time segmentation network, which is then evaluated on the real images from Cityscapes [5] (target domain).

- *Dataset*: GTA5 [5]
- *Training Set*: GTA5
- *Validation Set*: Cityscapes [5] validation split
- *Training epochs*: 50
- *Training resolution (GTA5)*: 1280x720
- *Test resolution (Cityscapes)*: 1024x512
- *Backbone*: ResNet18 (pre-trained on ImageNet) [2]
- *Semantic Classes*: 19
- *Metrics*: mIoU

Table 3) Domain Shift GTA5 → Cityscapes	**mIoU (%)**	**road**	**sidewalk**	**wall**	**fence**	**...**	**bicycle**
BiSeNet - 50 epochs							

How the performance change with respect to Table 2? Why?

b) Data augmentations to reduce the domain shift

A naive solution to improve the generalization capability of the segmentation network trained on the synthetic domain consists in the usage of data augmentations during training. Through them, we i) virtually expand the dataset size and ii) modify the visual appearance of source (synthetic) images in order to make them more similar to the target (real) ones.

Specifically, we repeat the previous experiment, introducing data augmentations at training time (e.g. horizontal flip, Gaussian Blur, Multiply, ecc.). The decision of what kind of algorithm is left to the student. Set the probability to perform augmentation to 0.5.

Table 4) Augmentations GTA5 → Cityscapes	**mIoU (%)**	**road**	**sidewalk**	**wall**	**fence**	**...**	**bicycle**
BiSeNet - 50 epochs - Aug. 1							
BiSeNet - 50 epochs - Aug. 2							

BiSeNet - 50 epochs - Aug. 1 + Aug. 2							
…							

4th STEP) DOMAIN ADAPTATION

To effectively tackle the problem of domain shift, various domain adaptation techniques has been proposed. Domain adaptation solutions are mainly divided into two approaches:

- **Adversarial approaches.** These methods involve a game between two parts of a model. One part, the feature extractor, tries to learn features that are indistinguishable between the source and target domains. The other part, the discriminator, tries to tell whether those features came from the source or target domain. This push-and-pull leads to features that are both discriminative (useful for the task) and domain-invariant (work well on both domains).
- **Image-to-image translation approaches.** The focus lies in translating images from the source domain to resemble the style of the target domain. The idea is that if we can make images from the different domains look similar, the model's performance will transfer more easily.

For the project extension, select a domain adaptation approach to implement.

FIRST CHOICE - Adversarial approach

You can assume:

- Source Synthetic Labelled Dataset: GTA5 [6]
- Target Real-World Unlabelled Dataset: Cityscapes [5]
- Implement discriminator function, like in [7]
- Take the best setting of step 3b (data augmentation) and perform training.

Table 5) Adversarial GTA5 → Cityscapes	**mIoU (%)**	**road**	**sidewalk**	**wall**	**fence**	**…**	**bicycle**
BiSeNet - 50 epochs							

SECOND CHOICE - Image-to-image approach

You can assume:

- Source Synthetic Labelled Dataset: GTA5 [6]
- Target Real-World Unlabelled Dataset: Cityscapes [4]
- Implement image-to-image adaptations: FDA [8] and DACS [9]
- Take the best setting of step 3b (data augmentation) and perform training.

Table 5) Image-to-image GTA5 → Cityscapes	**mIoU (%)**	**road**	**sidewalk**	**wall**	**fence**	**…**	**bicycle**
BiSeNet - FDA - 50 epochs							
BiSeNet - DACS - 50 epochs							

5th STEP) EXTENSION

The final step of the project involves exploring additional techniques or modifications to further improve the performance of the domain adaptation task. Here some examples:

- Apply Style Transfer Preprocessing: Preprocess source domain images with a style-transfer model to match the target domain's appearance.
- Explore alternative real-time networks (e.g. STDC PDF, PEM PDF, …).
- Explore Alternative Segmentation Losses: Investigate how using different losses impacts performance and domain adaptation (simpler alternative).
- Hyperparameter Tuning: Explore different learning rates, batch sizes, and data augmentation probabilities to optimize performance (simpler alternative).
- … use your imagination!

REFERENCES

[1] "A Brief Survey on Semantic Segmentation with Deep Learning", Shijie Hao, Yuan Zhou, Yanrong Guo, PDF

[2] "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFS", Liang-Chieh Chen, George Papandreou, Kevin Murphy, Alan L. Yuille PDF

[3] “BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation”, Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, Nong Sang, **PDF**
[4] “A Review of Single-Source Deep Unsupervised Visual Domain Adaptation”, Sicheng Zhao, Xiangyu Yue, Shanghang Zhang, Bo Li, Han Zhao, Bichen Wu, Ravi Krishna, Joseph E. Gonzalez, Alberto L. Sangiovanni-Vincentelli, Sanjit A. Seshia, Kurt Keutzer, **PDF**
[5] “The cityscapes dataset for semantic urban scene understanding.**”,** Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. **PDF**
[6] “Playing for data: Ground truth from computer games. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling” Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun, **PDF**
[7] “Learning to Adapt Structured Output Space for Semantic Segmentation”, Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, Manmohan Chandraker, **PDF**
[8] **“**FDA: Fourier Domain Adaptation for Semantic Segmentation” , Yanchao Yang, Stefano Soatto, **PDF**
[9] **“**DACS: Domain Adaptation via Cross-domain Mixed Sampling”, Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, Lennart Svensson, **PDF**