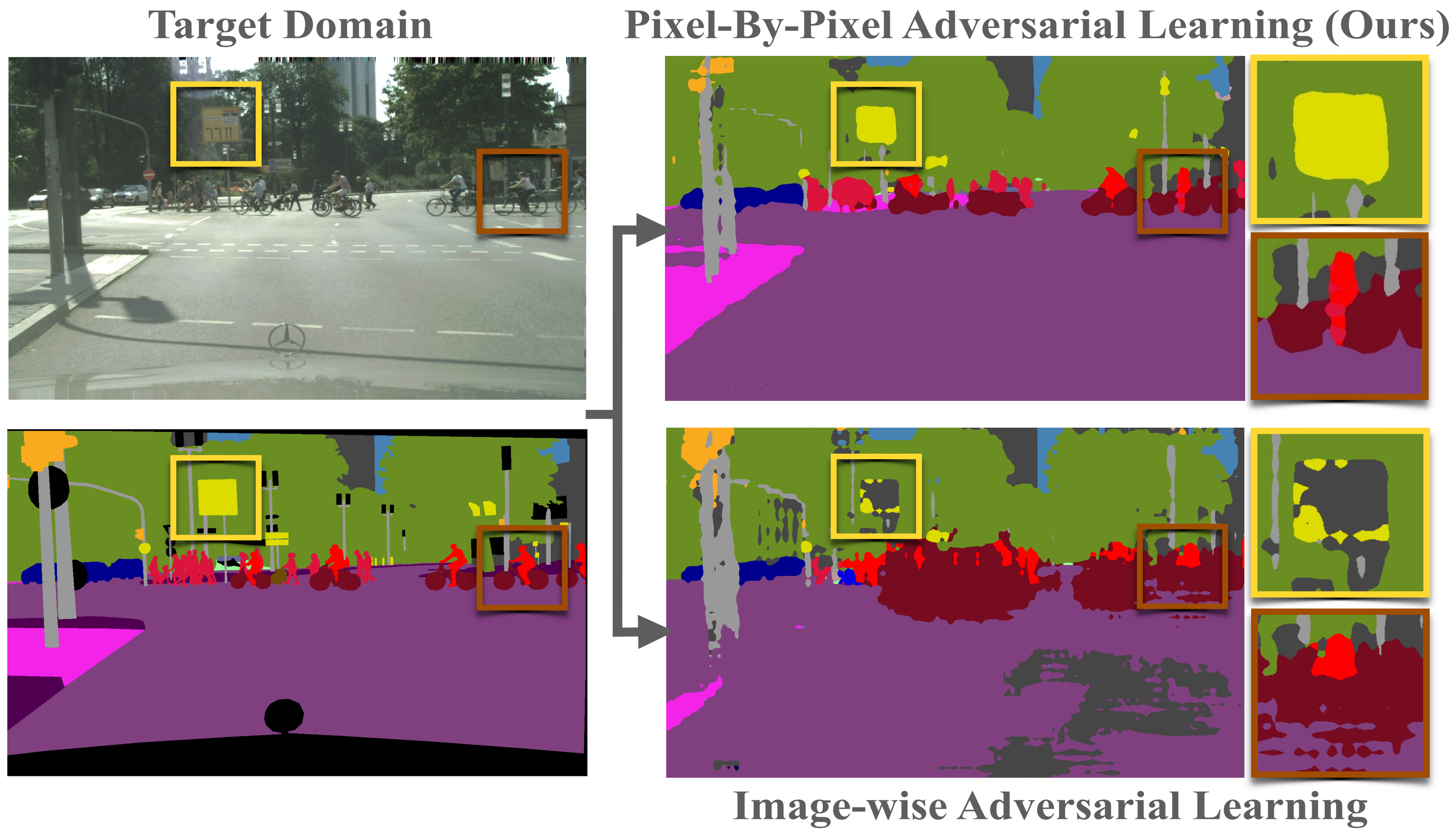
Politecnico di Torino
Dipartimento di Automatica e Informatica
DAUIN
ellis
European Laboratory for Learning and Intelligent Systems
ITALDESIGN

PhD in Computer and Control Engineering
XXXV cycle

Supervisor
Barbara Caputo

Robust and Fast Deep Semantic Segmentation

PhD Candidate: ***Antonio Tavera***



1. Setting

Semantic Segmentation: the task of assigning a semantic class to each pixel of an image. It is a foundamental task for some real-world applications, like autonomous driving.

Cross-Domain Few-Shot: training can use only few real-world annotated images (target) and many annotated synthetic images (source). Real-world data annotation is a costly operation.

One goal of my research is to teach the model to learn from a limited amount of labeled data.

2. Isssues

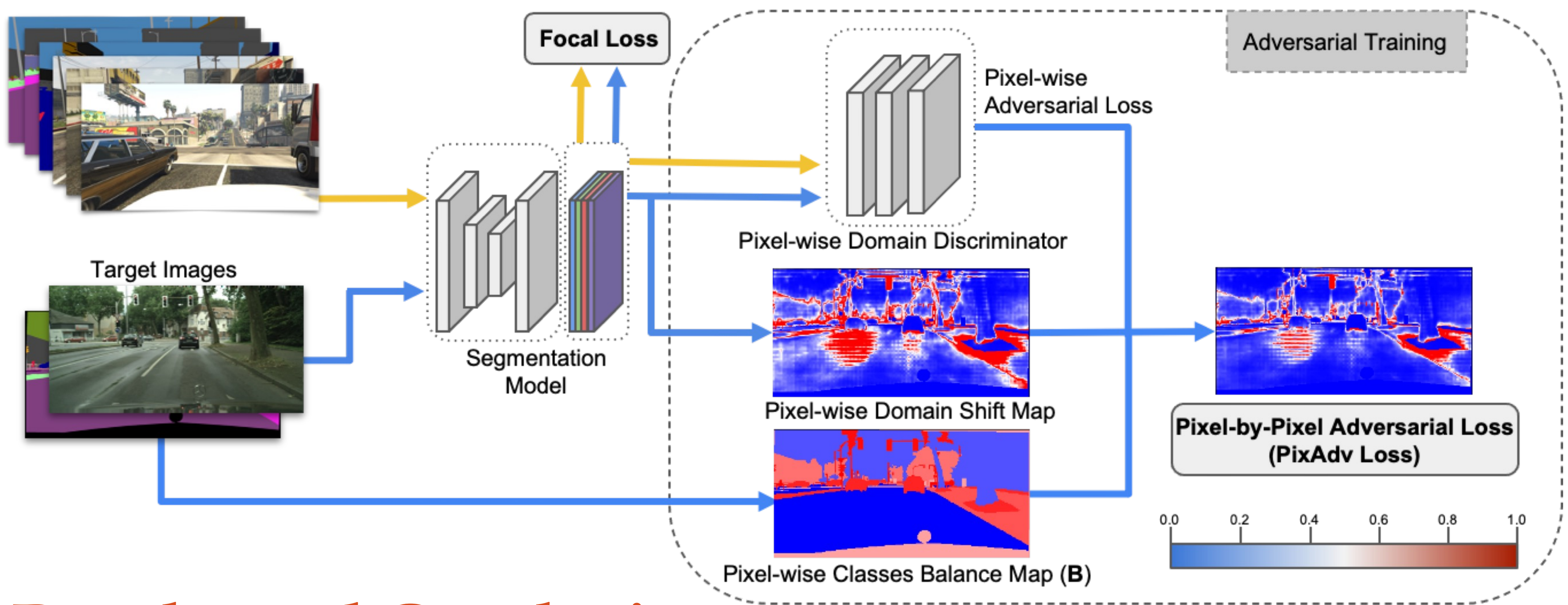
The **key challenges** arise from the inherent imbalance between the source and target domains, which is caused by: (i) Visual domain shift; (ii) Different number of images; (iii) Pixel-wise imbalance among segmented classes.

3. Method

We introduce a Pixel-by-Pixel adversarial learning framework (**PixDA**[1]) to prioritize and improve pixel alignment using three criteria: (i) align the source and target domain, (ii) avoid to further align correctly represented pixel, and (iii) regularize the training of infrequent classes. The new pixel-wise domain discriminator collaborates with our new PixAdv adversarial loss (**PixAdv**) to align each pixel according to a combination of two different terms:

1[st] term) related to the network classification confidence.

2[nd] term) represents the imbalance of the pixels.

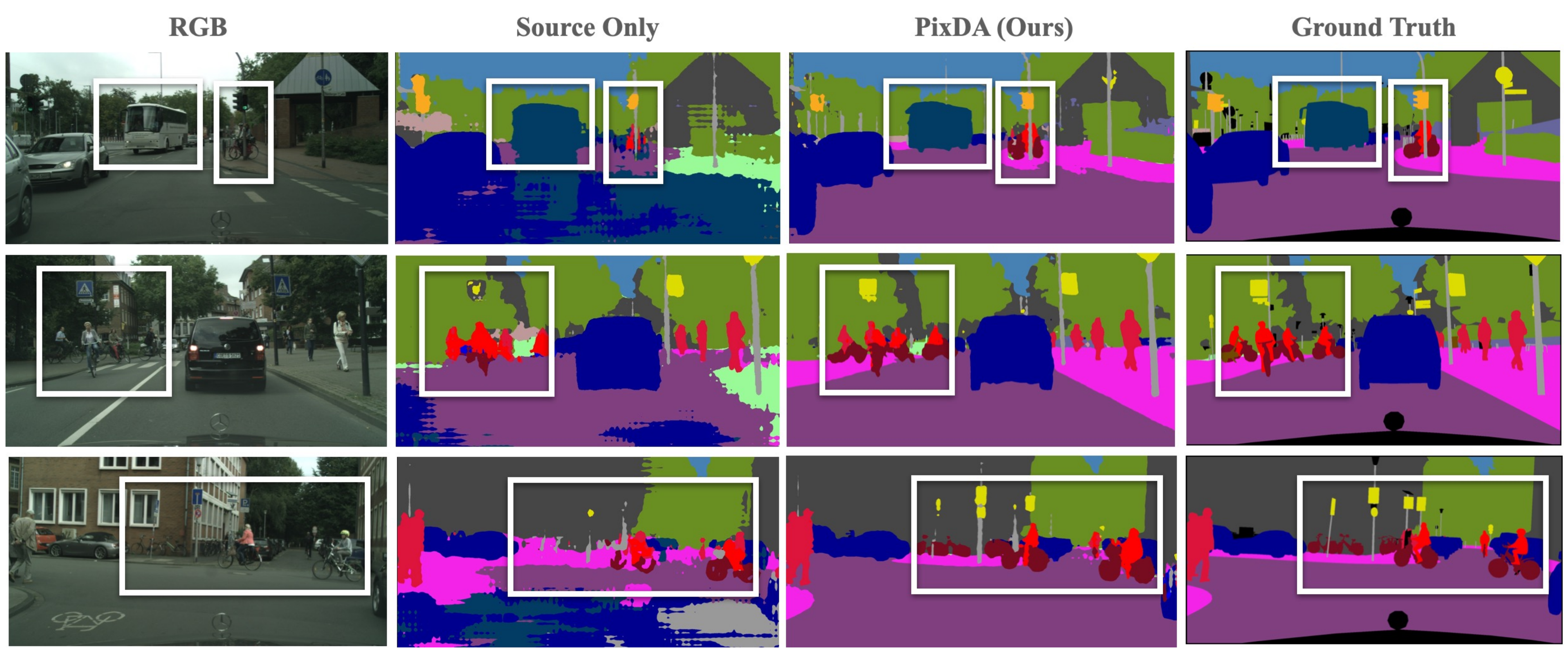


4. Results and Conclusion

Shot	Method	mIoU[19]	mIoU[Well]	mIoU[Under]
	Source Only	30.7	41.4	20.1
1 shot	NAAE	27.6	42.3	7.3
	Joint Training	31.3	41.2	17.7
	FDA	42.0	54.0	25.6
	Fine-Tuning	45.3	56.0	26.8
	FSDA	46.4	60.4	27.2
	PixDA	**51.2**	**63.3**	**34.5**
3 shot	NAAE	34.7	49.3	13.7
	Joint Training	34.5	44.7	20.6
	FDA	41.8	53.5	25.7
	Fine-Tuning	45.6	62.7	31.6
	FSDA	51.0	63.9	33.3
	PixDA	**54.5**	**65.1**	**40.1**
5 shot	NAAE	40.0	53.9	20.9
	Joint Training	34.8	46.0	19.5
	FDA	41.0	54.4	22.7
	Fine-Tuning	53.5	64.1	38.9
	FSDA	53.6	63.9	38.2
	PixDA	**55.6**	**65.3**	**42.1**

The table compares several methods for few-shot domain adaptation in terms of mean Intersection over Union (mIoU) on the standard **GTA to Cityscapes** protocol. Our method achieves **state-of-the-art results** in all the **1-to-5 shot** settings.

The introduction of our loss outperforms by a large margin the standard image-wise adversarial training, confirming that weighting each pixel contribution is advantageous to prevent negative transfer and overfitting.



[1] **Tavera A**. *et al.*, Pixel-by-Pixel Cross-Domain Alignment for Few-Shot Semantic Segmentation, IEEE Winter Conference on Applications of Computer Vision (WACV) 2022.

For more details

antonio.tavera@polito.it
https://taveraantonio.github.io