University of Information Technology, VNU-HCM
Faculty of Computer Science

SEMANTIC IMAGE SEGMENTATION IN THE DARK WITH DOMAIN ADAPTATION METHOD

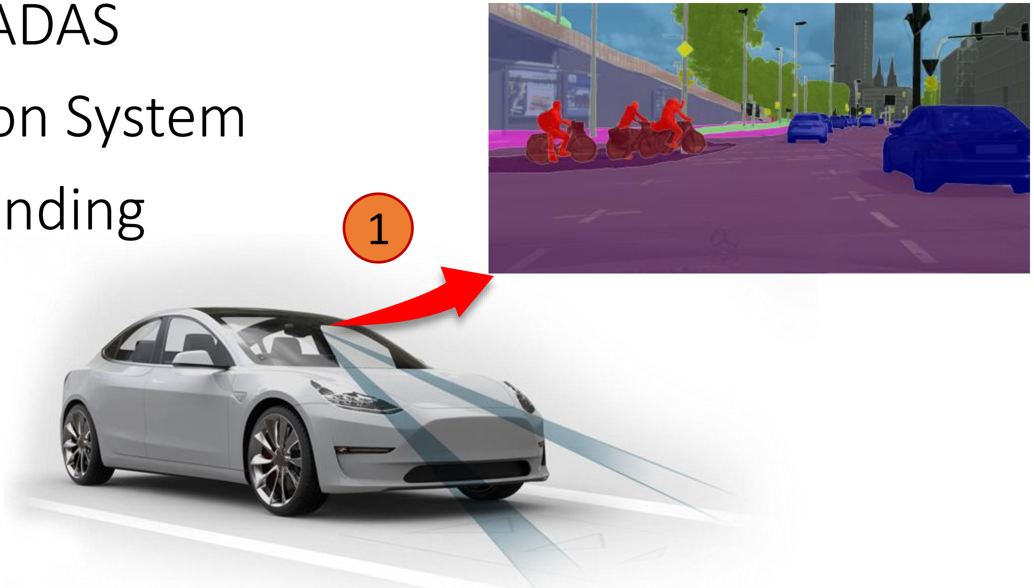
THESIS PRESENTATION

Nguyen Thanh Danh – 17520324
Phan Nguyen – 17520828
Advisor: Dr. Nguyen Vinh Tiep

Contents

Practical Context

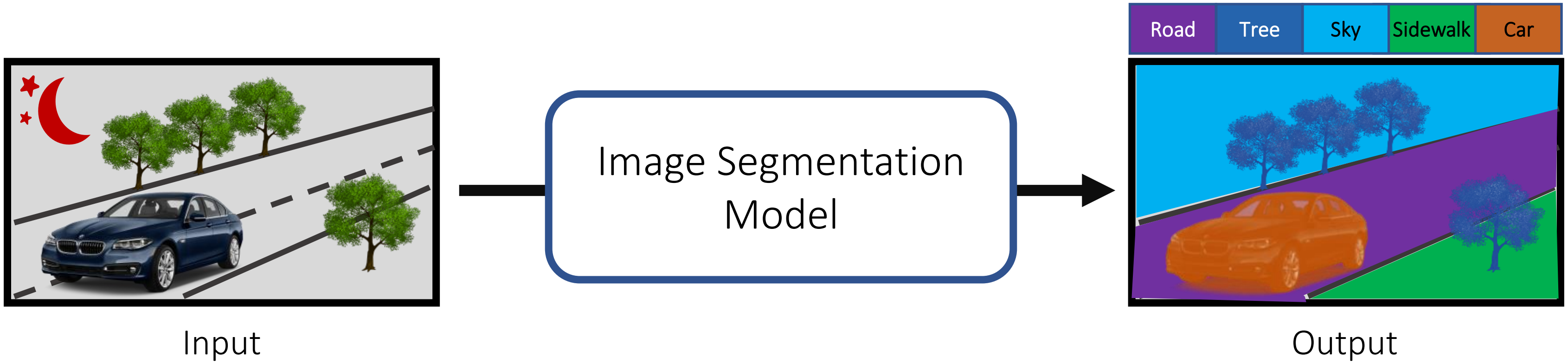
1. Autonomous Vehicles - ADAS
2. Medical Recommendation System
3. Satellite Image Understanding



Problem Definition

Semantic image segmentation on *nighttime* cityscapes images

- **Input**: nighttime cityscapes images
- **Output**: segmentation maps



Challenges

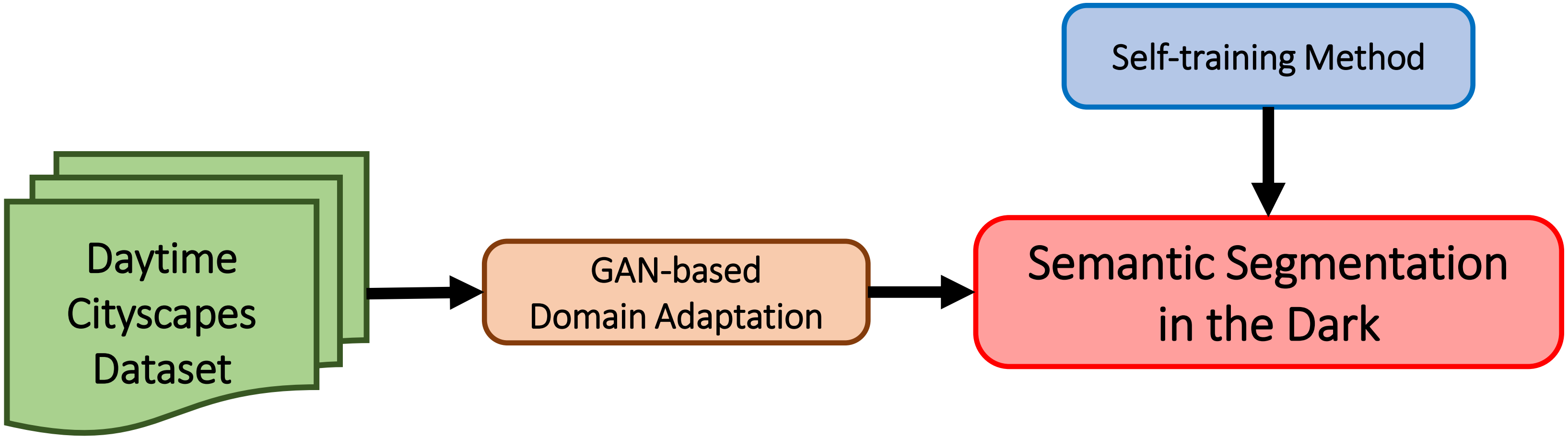
- **Lack of annotated dataset** for nighttime cityscapes segmentation
- External conditions: light blur, rainy, etc.

Exemplary nighttime images

Domain Adaptation Method

Objectives

Solve **semantic image segmentation in the dark**
with **GAN-based domain adaptation method**
to leverage existing daytime cityscapes dataset
along with **self-training method**



Our Contributions

1. Propose a framework for **semantic image segmentation in the dark** with **domain adaptation method**
2. Propose **a loss function** for semantic image segmentation
3. Build a **nighttime cityscapes dataset** with GAN

Related Work

1. Dark Model Adaptation. ITSC2018

Daytime → Twilight → Nighttime

Night-Model Segmentation

2. See clearer at night. ISOP2019

Daytime — Day-Model Segmentation

Nighttime — Night-Model Segmentation

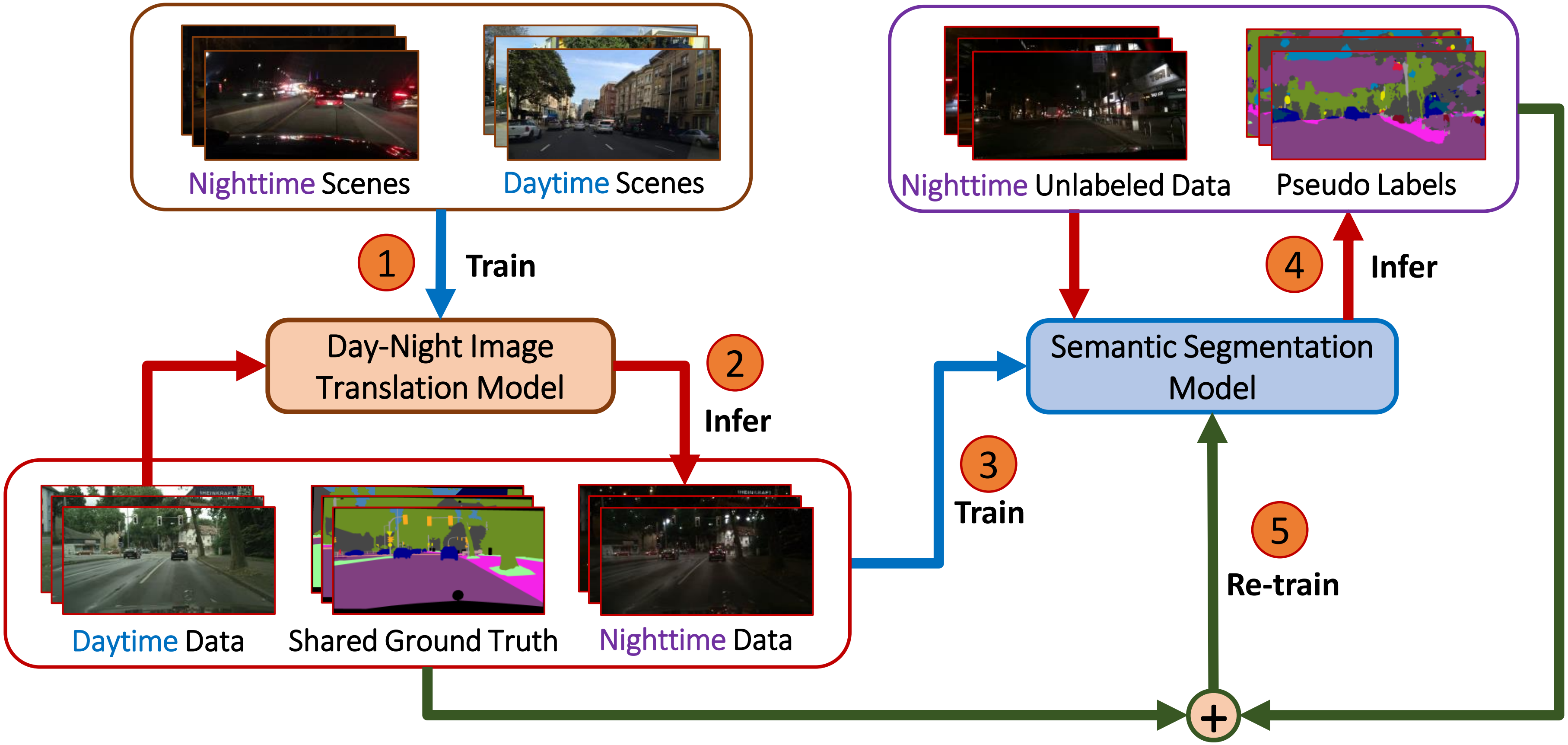
3. Self-training. NIPS2020

Unlabeled Data

Refine

Trained Model

Proposed Framework



Dataset

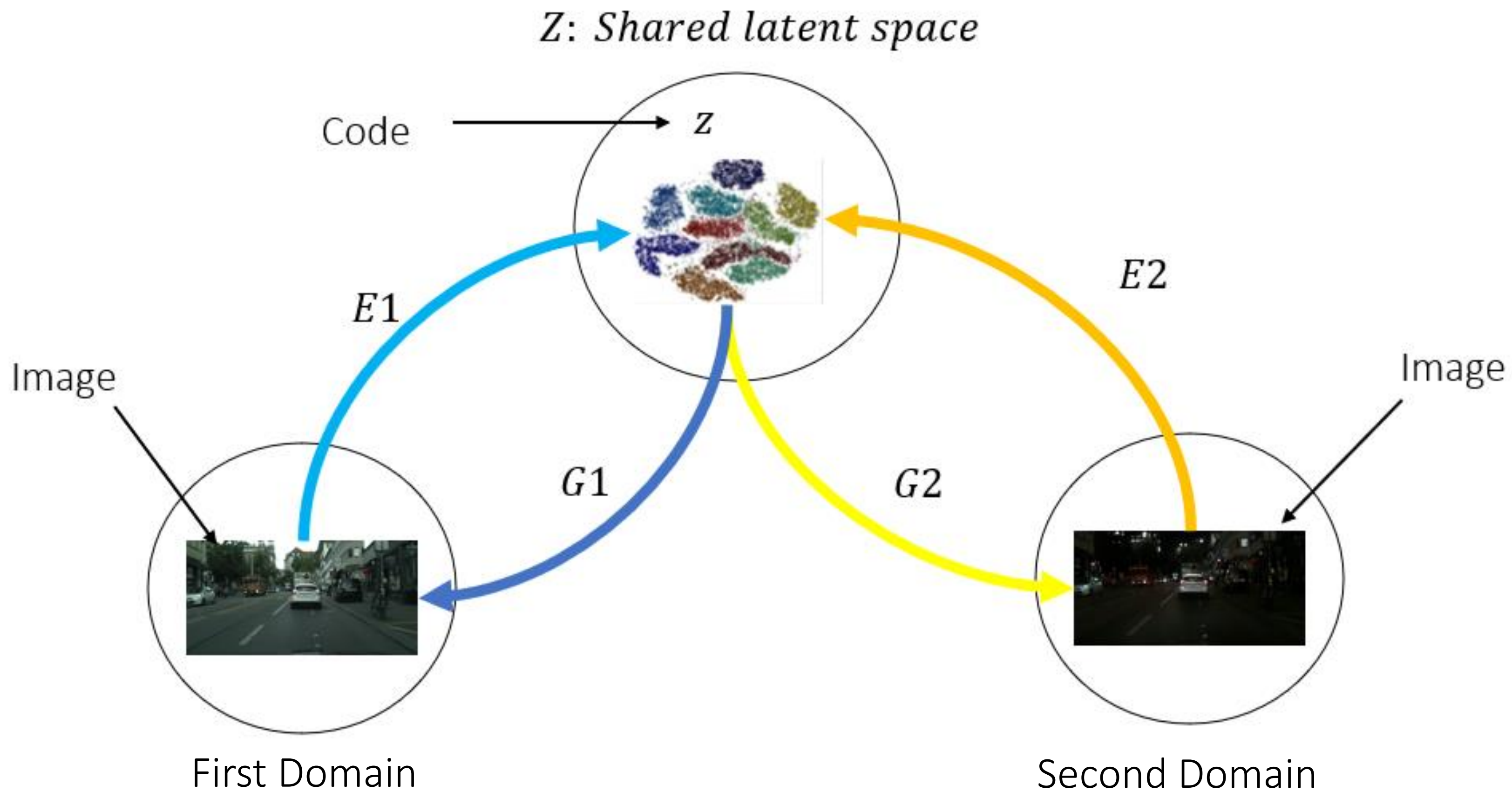
- **NEXET Dataset**: ~50k day, night, twilight images
- **Histogram-based method** to separate images into 2 domains: daytime and nighttime (ignore twilight)

19,858 Daytime Images

19,523 Nighttime Images

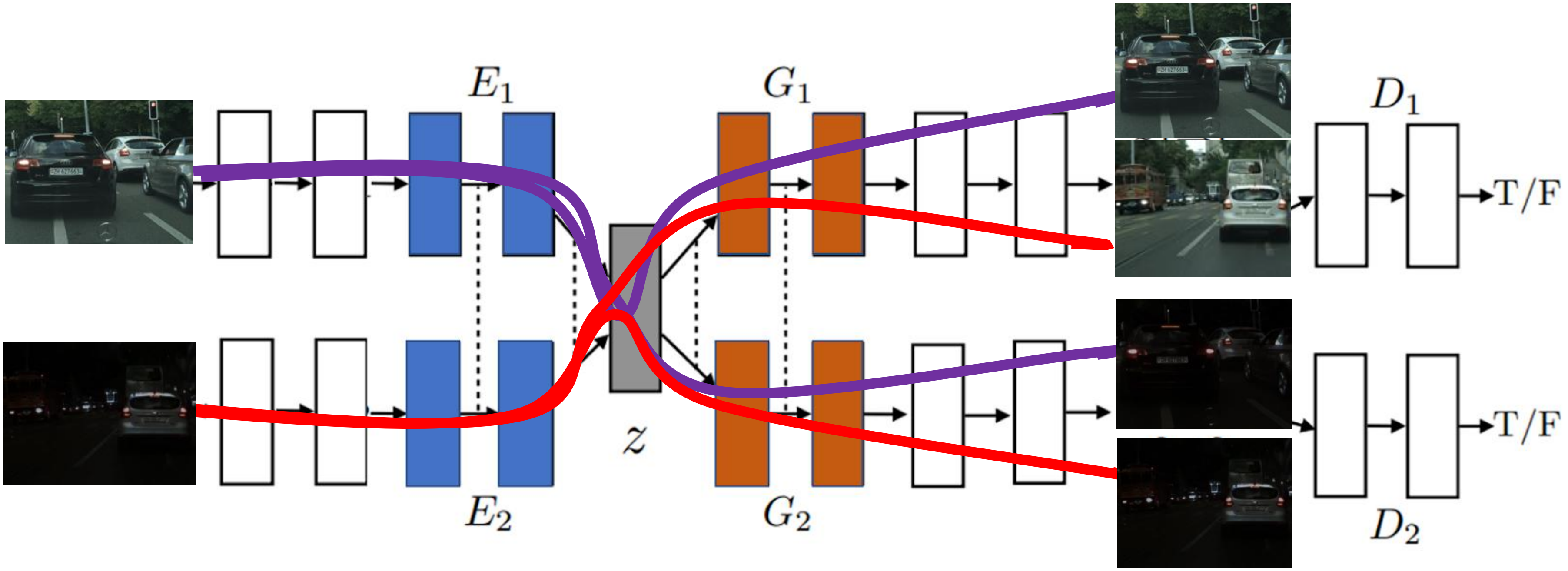
GAN-based Method

Assumption: **shared latent space**



GAN-based Method

1. Variational Autoencoders (VAEs)
2. Weight-sharing
3. GAN



Day2Night Translation Results

- **Mismatch** vehicle/traffic lights
- **Correctly match** the lights **(w/ Perceptual Loss)**

Original Images | Initial Results | w/ Perceptual Results

Quantitative Results

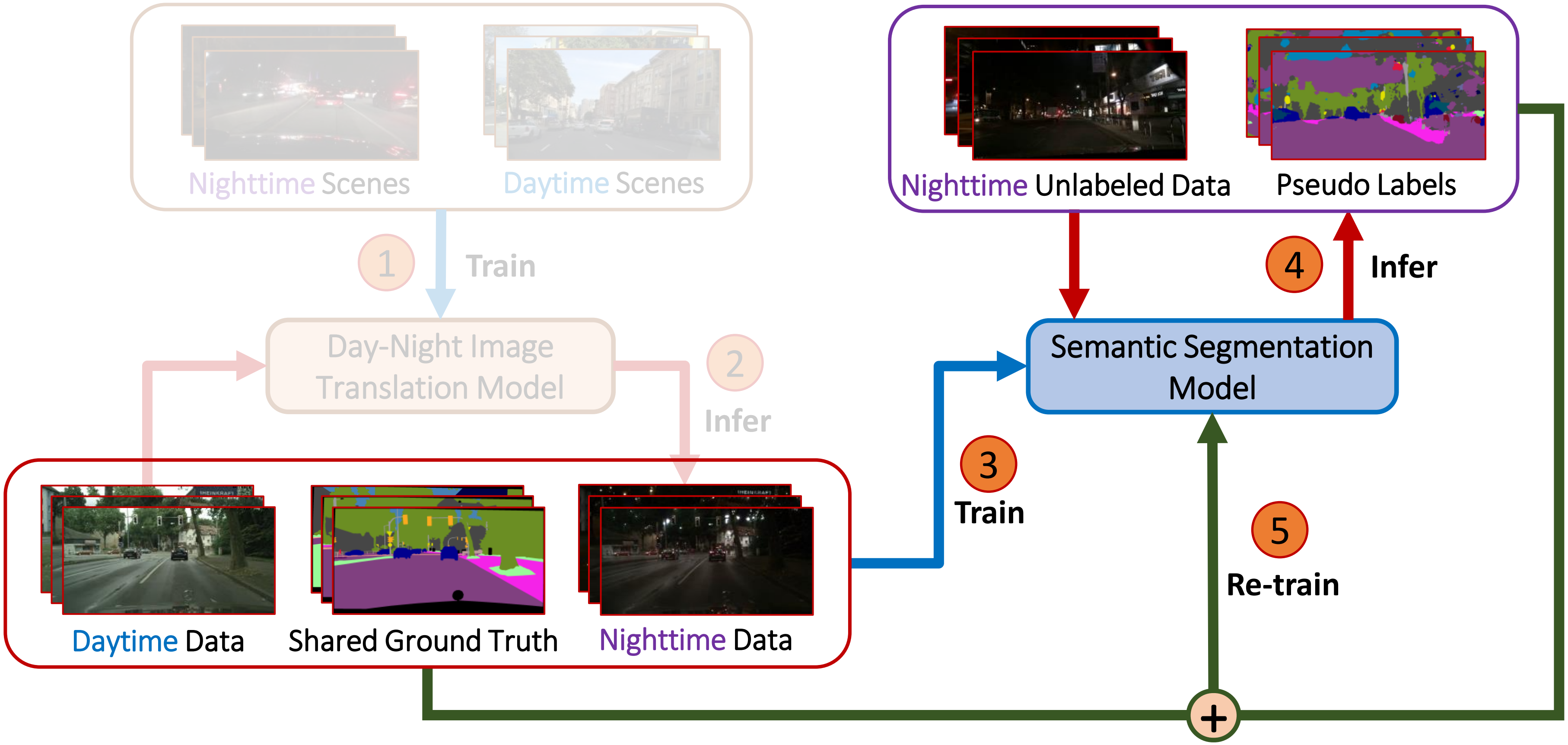
$$\boldsymbol{FID} = ||\mu_1 - \mu_2||^2 + Tr(C_1 + C_2 - 2\sqrt{C_1 C_2})$$

- μ: mean
- C: Covariance

FID score shows the differences of **generated** and **real** images.

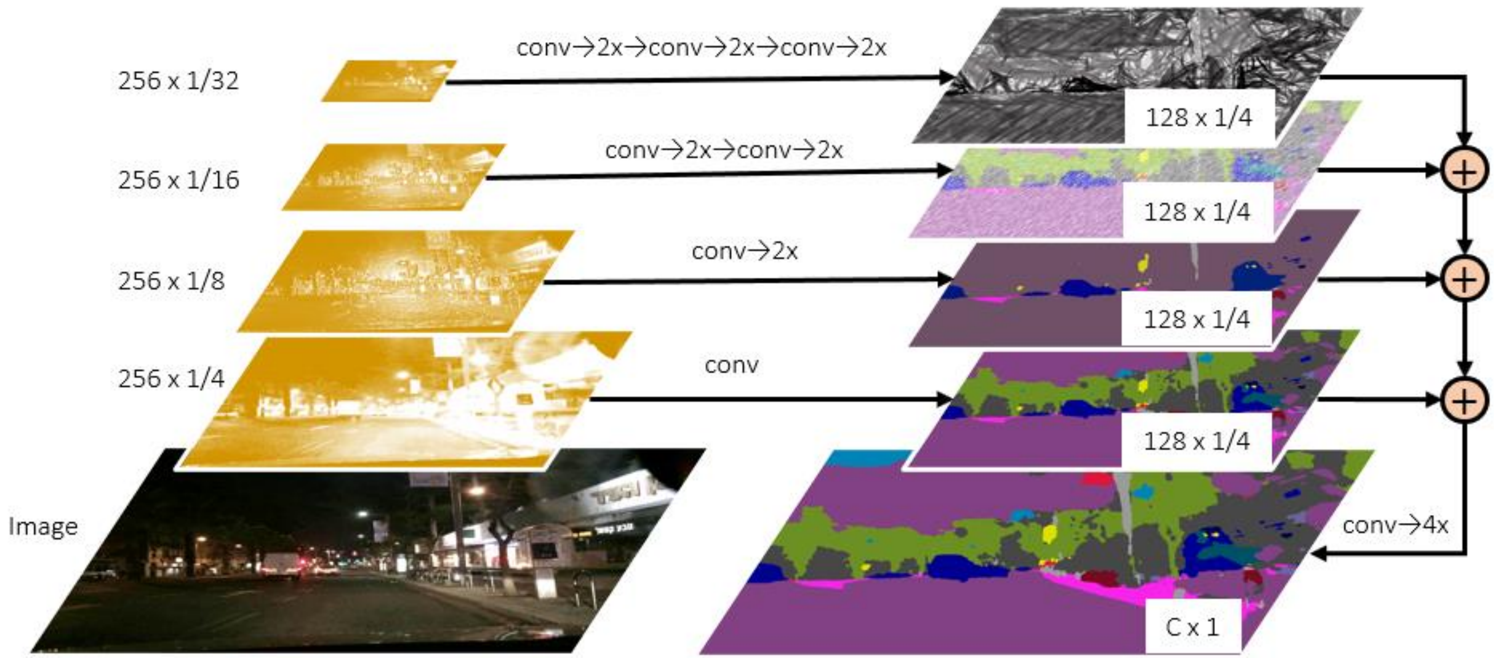
ID	Method	FID_night
1	UNIT w/o Perceptual	98.39
2	UNIT w Perceptual	97.68

Semantic Segmentation Component



Semantic Segmentation Model

- Panoptic Feature Pyramid Networks – ResNet101
- Specifications:
  - Ensemble **low and high level features**
  - Extract **multi-scale features**



Panoptic Feature Pyramid Networks

Proposed Combined Loss

- Measure **differences** among **couple of pixels**

$$L_{pixel}(p_t) = Cross\_Entropy\_Loss(p_t) = -\log(p_t)$$

- Solve **imbalanced problem** of major classes

$$L_{balance}(p_t) = Focal\_Loss(p_t) = -\alpha_t\,(1 - p_t)^{\gamma} \cdot log(p_t)$$

$$p_t = \begin{cases} p(x_i) & if\ y = 1 \\ 1 - p(x_i) & if\ y = 0 \end{cases}$$

- We propose:

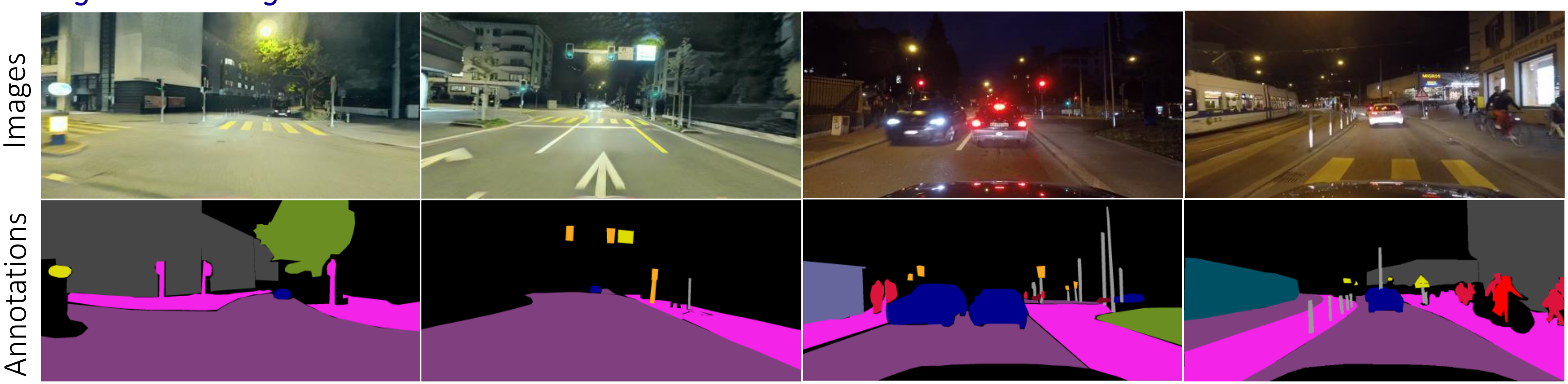
$$\boxed{L_{combined}(p_t) = \alpha\, L_{pixel}(p_t) + (1 - \alpha) L_{balance}(p_t)}$$

$(Weight\ \alpha = 0.5)$

Segmentation Dataset

Cityscapes. M. Cordts et al. CVPR2016

Nighttime Driving Test. D.Dai and L. Gool. ITSC2018

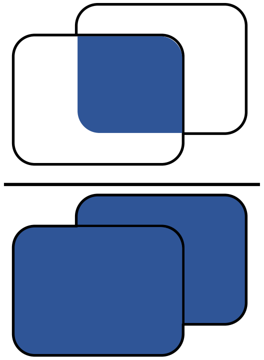


Void	Road	Sidewalk	Building	Wall	Fence	Pole	Traffic Light	Traffic Sign	Vegetation
Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motorcycle	Bicycle

Evaluation Metrics

- Pixel Accuracy (PA)
- Class Accuracy (CA)
- **Mean Intersection over Union (mIoU)**
- Frequency Weighted Intersection over Union (FWIoU)

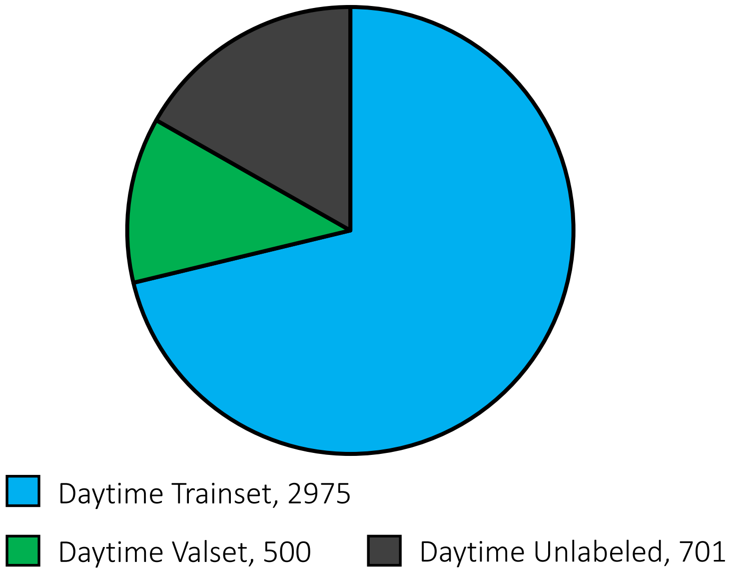
$$IoU = \frac{Intersection}{Union} =$$



Experiment 1

- Daytime Cityscapes images
- Self-train with daytime **CamVid** dataset

Data Distribution



➔ Self-training from a checkpoint is better than from scratch

ID	Configuration	mIoU
1.1	FPN-res101-daytime-Cityscapes	27.5
1.2	FPN-res101-self-training-from-scratch	27.1 (-0.4)
1.3	FPN-res101-self-training-from-ckpt	29.0 (+1.5)

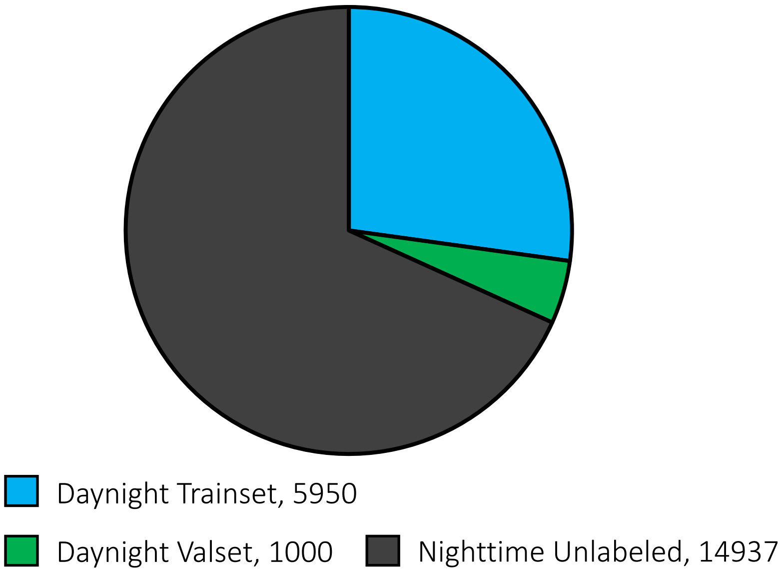
Experiment 2

- Day-night Cityscapes images
- Self-train with 14.937 nighttime images

➔ Minimizing image domain distance improves model performance

➔ Self-training is not useful?

Data Distribution



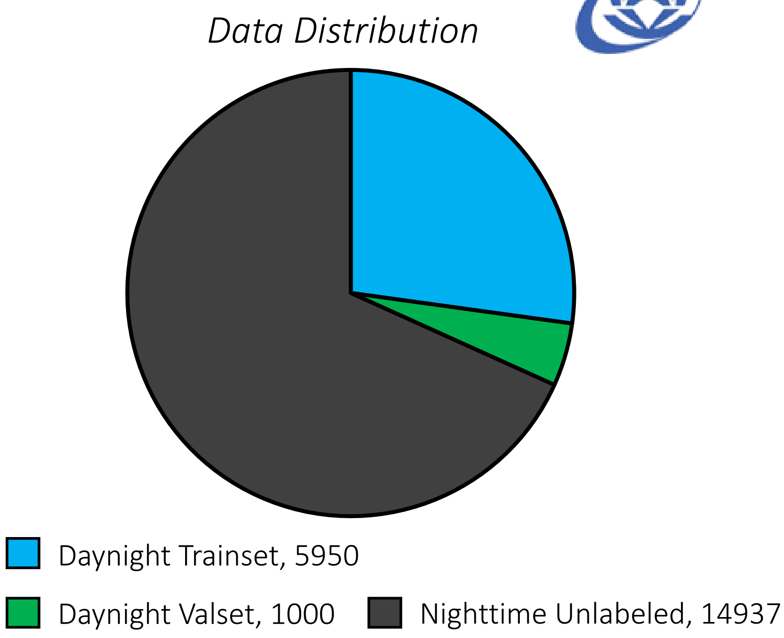
ID	Testset	Configuration	mIoU
2.1	Origin	FPN-res101-daynight	31.5 (+2.5)
2.2		FPN-res101-self-training-15k-from-ckpt-2.1	28.8 (-2.7)
2.3	Converted	FPN-res101-daynight	25.2
2.4		FPN-res101-self-training-15k-from-ckpt-2.1	24.7 (-0.5)

Experiment 3

- **Day-night** Cityscapes images
- Self-train with **14.937 nighttime images**
- Image Translation with perceptual loss

→ Perceptual loss maintains semantic features when translating images

→ Self-training is not useful?



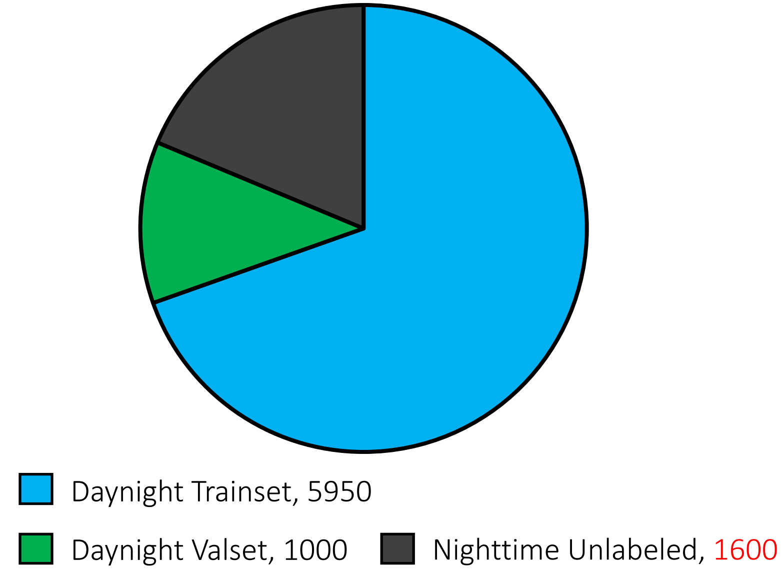
ID	Testset	Configuration	mIoU
3.1	Origin	FPN-res101-daynight	33.9 (+2.4)
3.2		FPN-res101-self-training-15k-from-ckpt-3.1	32.1 (-1.8)
3.3	Converted	FPN-res101-daynight	29.3
3.4		FPN-res101-self-training-15k-from-ckpt-3.1	28.4 (-0.9)

Experiment 4

- **Day-night** Cityscapes images
- Self-train with 1.600 **nighttime images** (based on histogram)
- Image Translation with **perceptual loss**

➔ Self-training is useful with suitable amount of unlabeled data

Data Distribution



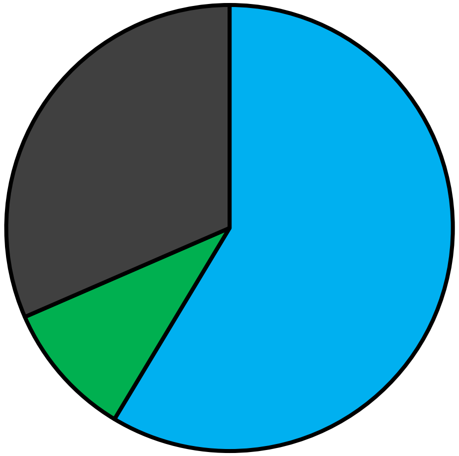
ID	Testset	Configuration	mIoU
3.1	Origin	FPN-res101-daynight	33.9
4.1		FPN-res101-self-training-1k6-HIS-ckpt-3.1	34.2 (+0.3)
3.3	Converted	FPN-res101-daynight	29.3
4.2		FPN-res101-self-training-1k6-HIS-ckpt-3.1	29.8 (+0.5)

Experiment 5

- Only-night Cityscapes images
- Self-train with **1.600 nighttime images** (based on **histogram**)
- Image Translation with **perceptual loss**

➔ An extra training on the target prediction domain improves model performance

Data Distribution

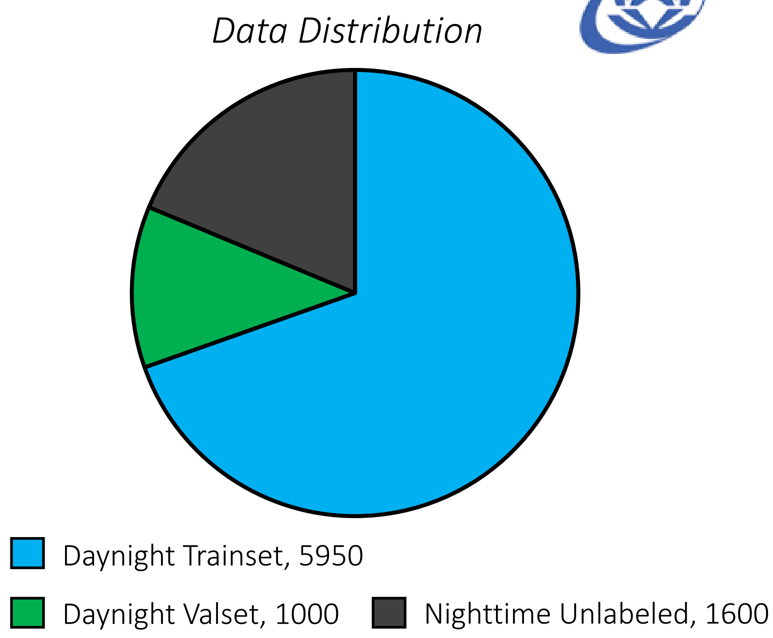


ID	Configuration	mIoU
5.1	FPN-res101-onlynight	29.6
5.2	FPN-res101-**morenight**-from-ckpt-3.1	**34.7 (+0.8)**
5.3	FPN-res101-self-training-1k6-HIS-from-ckpt-5.1	29.8 (+0.2)
5.4	FPN-res101-self-training-1k6-HIS-from-ckpt-5.2	33.3 (-1.4)

Experiment 6

- **Day-night** Cityscapes images
- Self-train with **1.600 nighttime images** (based on **histogram**)
- Image Translation with **perceptual loss**
- Segmentation with **Focal Loss**

➔ Focal loss result is not higher than cross entropy loss

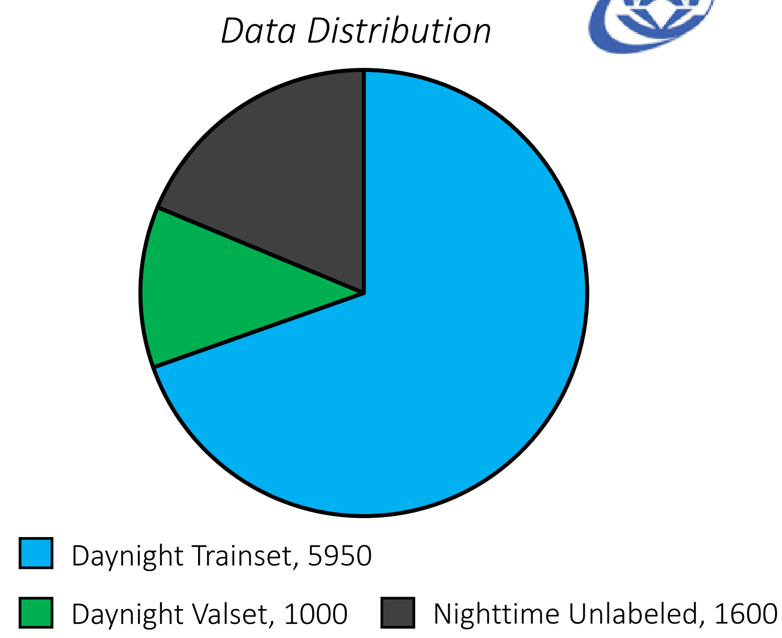


ID	Configuration	mIoU
3.1	FPN-res101-daynight-CE	33.9
4.1	FPN-res101-self-training-1k6-HIS-from-ckpt-3.1-CE	**34.2** (+0.3)
6.1	FPN-res101-daynight-FL	26.9
6.2	FPN-res101-self-training-1k6-HIS-from-ckpt-6.1-FL	28.3 **(+1.4)**

Experiment 7

- **Day-night** Cityscapes images
- Self-train with **1.600 nighttime images** (based on FID)
- Image Translation with **perceptual loss**
- Segmentation with Proposed Combined Loss (CL)

➔ FID method helps choose similar domain images

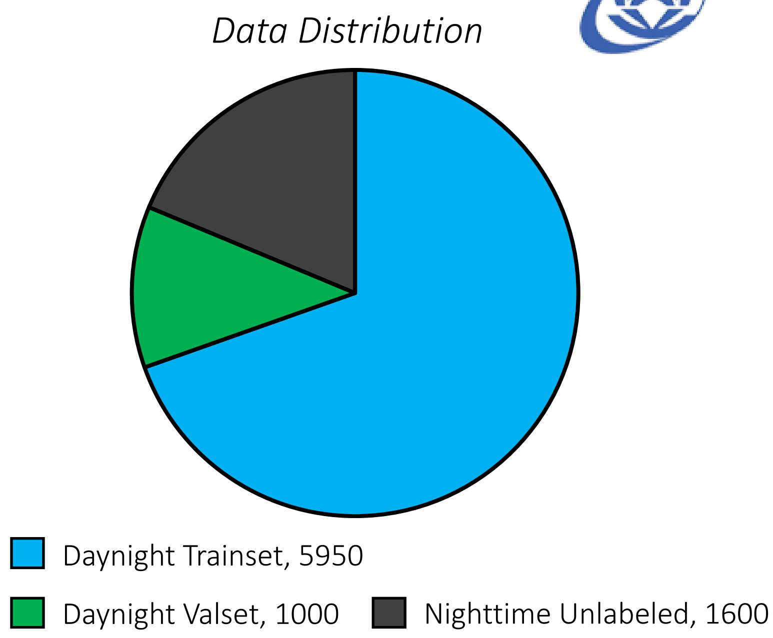


ID	Configuration	mIoU
3.1	FPN-res101-daynight-CE	33.9
7.1	FPN-res101-self-training-1k6-**FID**-from-ckpt-3.1-CE	38.8 (+4.9)
7.2	FPN-res101-self-training-1k6-**FID**-from-ckpt-3.1-CL	39.3 (+5.4)
7.3	FPN-res101-self-training-1k6-**HIS**-from-ckpt-3.1-CL	33.1 (-0.8)

Experiment 8

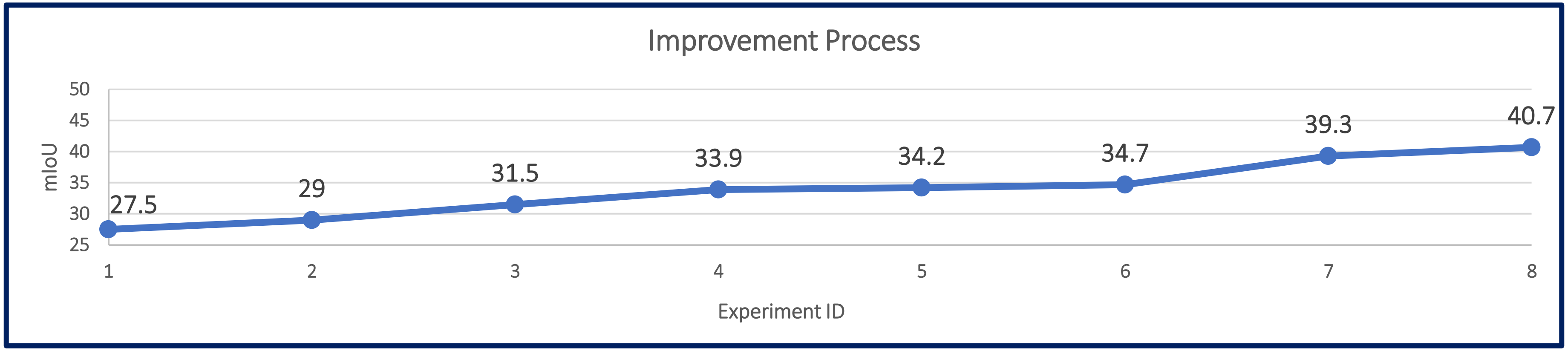
- **Day-night** Cityscapes images
- Self-train with **1.600 nighttime images** (based on **FID**)
- Image Translation with **perceptual loss**
- Segmentation with **Proposed Combined Loss (CL)**

➔ Our total configuration achieves the finest performance



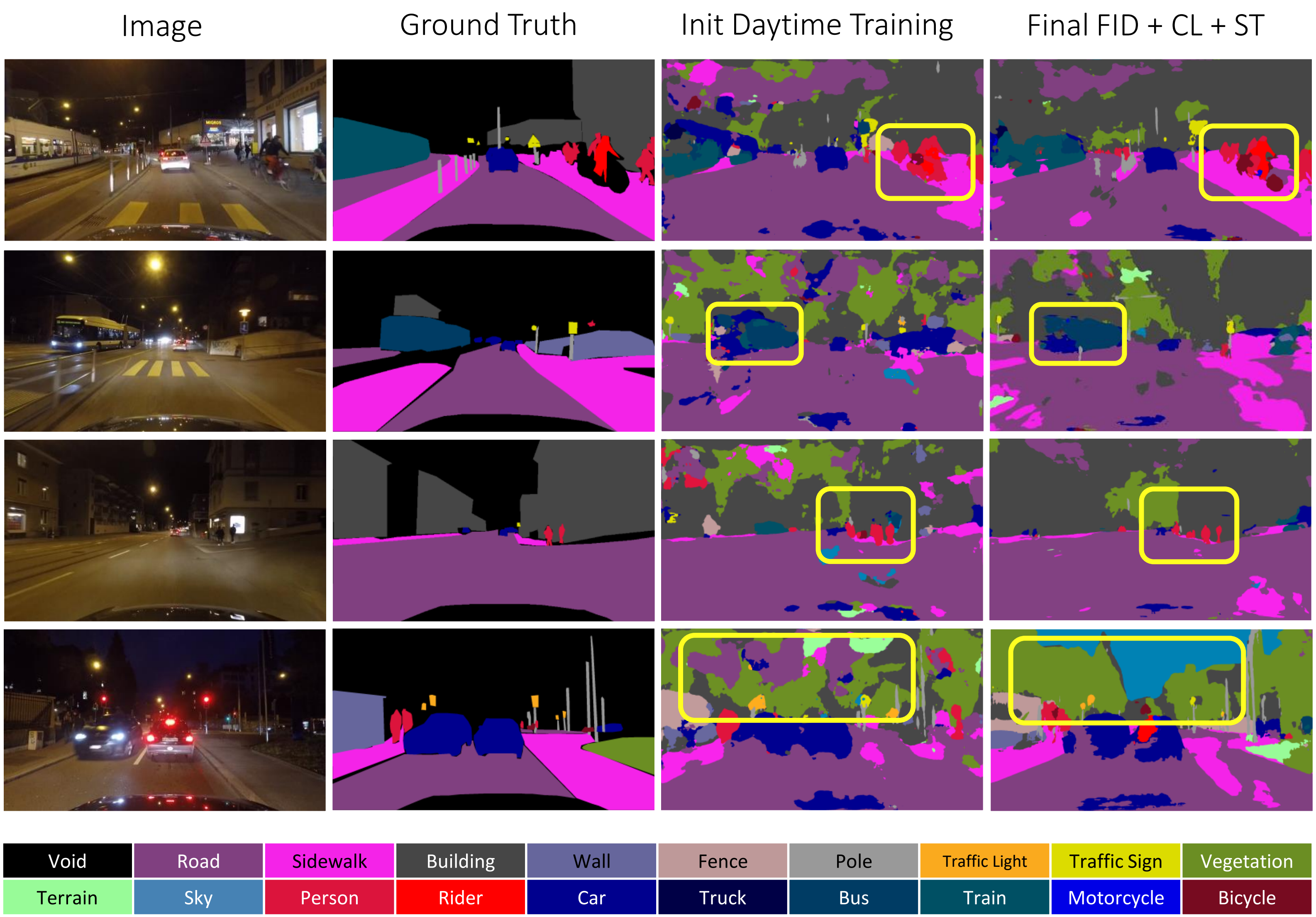
ID	Configuration	mIoU
3.1	FPN-res101-daynight-CE	33.9
5.2	**FPN-res101-morenight-from-ckpt-3.1**	34.7 (+0.8)
8.1	FPN-res101-self-training-1k6-HIS-from-ckpt-5.2-CE	37.8 (+3.9)
8.2	FPN-res101-self-training-1k6-**FID**-from-ckpt-8.1-CE	39.5 (+5.6)
8.3	FPN-res101-self-training-1k6-**FID**-from-ckpt-8.1-CL	40.7 (+6.8)

Improvement Process



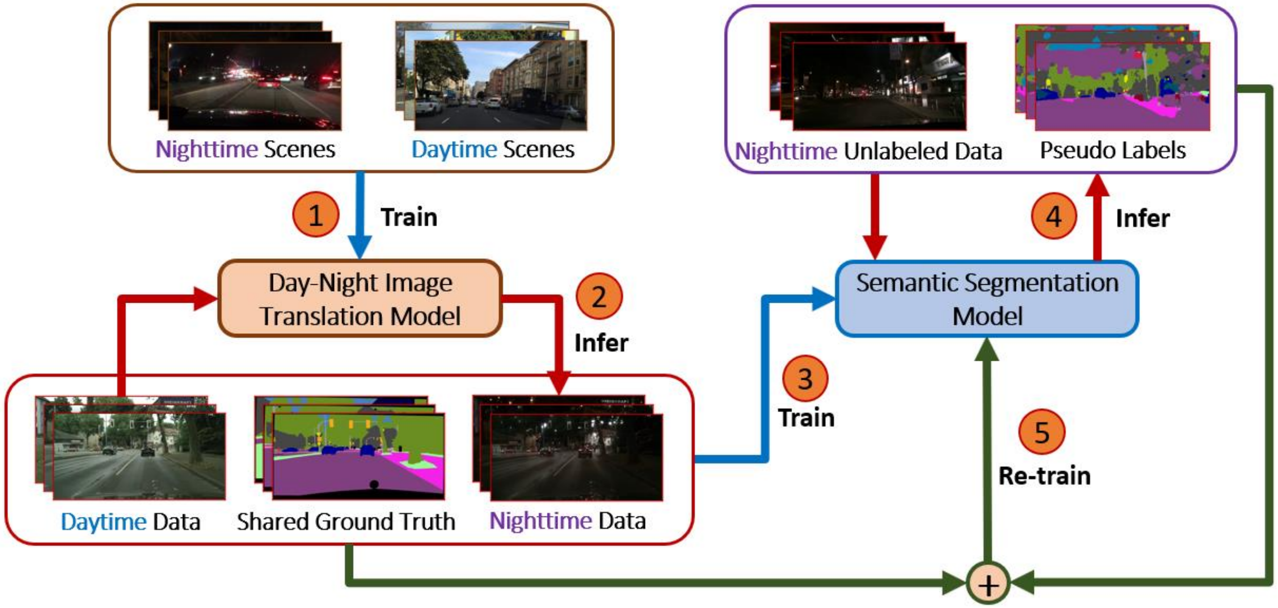
ID	Configuration	mIoU
1	Daytime Cityscapes dataset	27.5
2	Self-train with daytime data	29.0
3	Train and Self-train with day-night data	31.5
4	Add perceptual loss to translate images	33.9
5	Use histogram-based method to choose extra data	34.2
6	Refine day-night model with more nighttime images	34.7
7	Use FID to choose extra data and combined loss	39.3
8	Total FID, combined loss, self-training from morenight ckpt	40.7

Experiments Visualization



Summary – Our Contributions

1. Propose a framework for **semantic image segmentation in the dark** with **domain adaptation method**
2. Propose **a loss function** for semantic image segmentation
3. Build a **nighttime cityscapes dataset** with GAN



Publication

- Xuan-Duong Nguyen, Anh-Khoa Nguyen Vu, **Thanh-Danh Nguyen**, **Nguyen Phan**, Bao-Duy Duyen Dinh, Nhat-Duy Nguyen, Tam V. Nguyen, Vinh-Tiep Nguyen, Duy-Dinh Le: *Adaptive Detection-Tracking-Counting Framework for Multi-Vehicle Motion Counting, Image and Vision Computing – IMAVIS (**ISI Q1**), 2021. (under review)*

Thanks for your attention!