



**Politecnico  
di Torino**

# KNOWLEDGE DISTILLATION FOR SEMANTIC SEGMENTATION APPLIED TO AUTONOMOUS DRIVING IN ADVERSE WEATHER CONDITION

Supervisors:

Prof. Paolo GARZA

Dott. Edoardo ARNAUDO

Dott. Marco GALATOLA

Dott. Stefano BERGIA

Candidate:

Federico BUSSOLINO

# Thesis objectives and motivation

The objective of this thesis is to assess performances of semantic segmentation algorithm that can run on edge devices, using as data in input an RGB monocular camera, in order to reduce costs of deployment of autonomous driving systems.

The experiments are tested on adverse weather condition dataset since current autonomous driving systems cannot reach SAE 5 level of autonomy, which requires navigation in adverse weather.

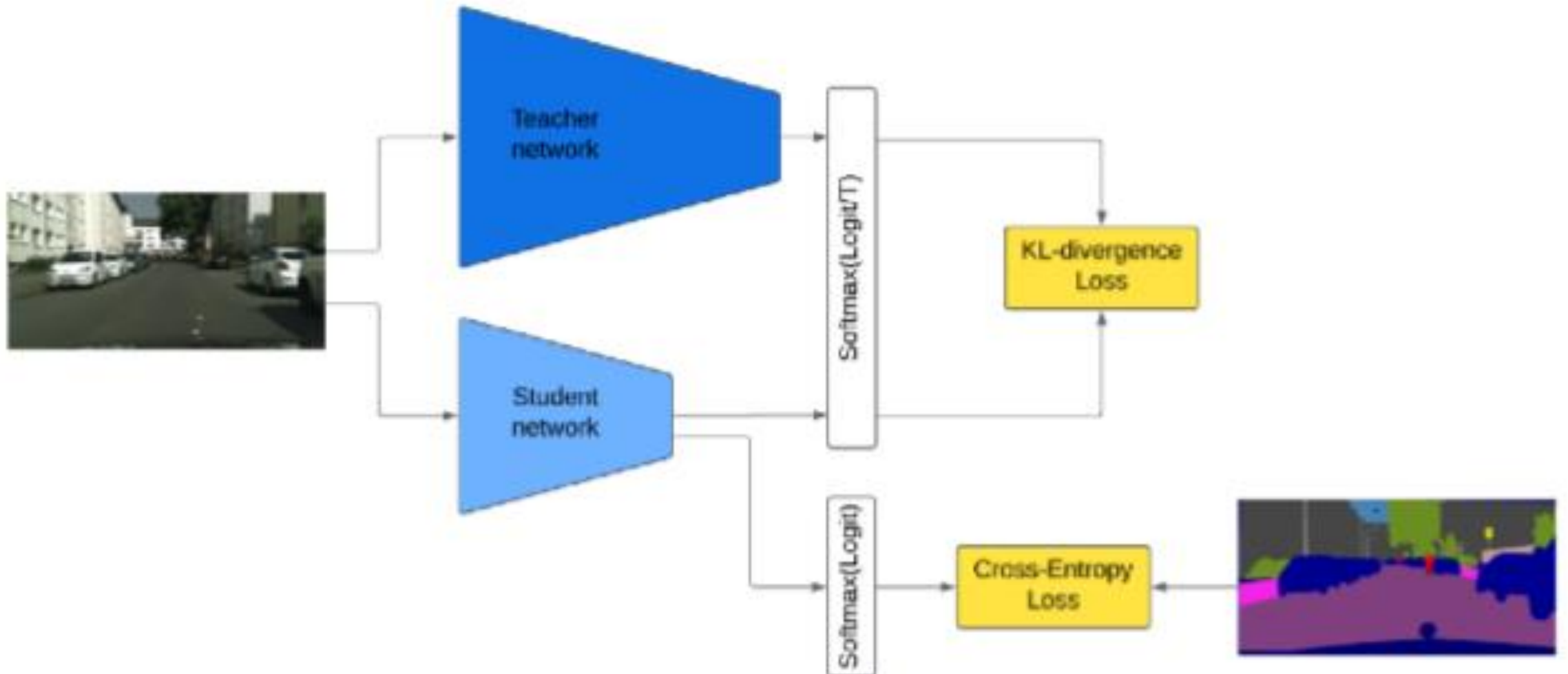
# Semantic segmentation



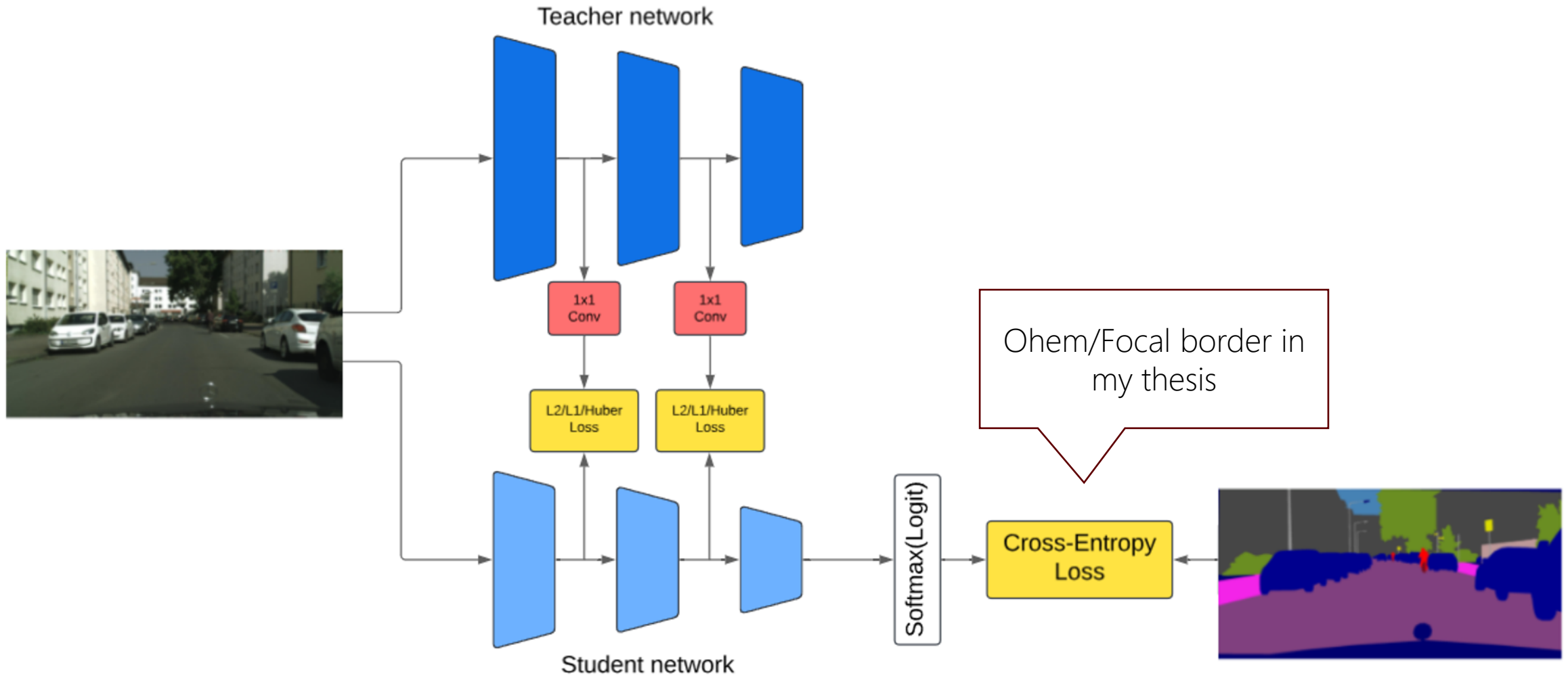
ROAD SIDEWALK BUILDING WALL FENCE POLE TRAFFIC SIGN VEGETATION TERRAIN SKY RIDER CAR BIKE BACKGROUD

- Use CNN or ViT to extract feature at different scales
- Aggregate feature maps of different scales
- $\# \text{Logits} = \# \text{Pixels}$

# Knowledge distillation based on logits (Hinton, 2015)



# Knowledge distillation based on feature maps (FitNets, 2014)



# Datasets

- ACDC (Adverse conditions with correspondence) → Used to train
- Cityscapes → Used to select baselines, standard benchmark for semantic segmentation, can be used to expand training data
- BDD100K → Widely used as a benchmark, useful for multitask benchmark, can be used to expand training data

## Baselines (real-time networks)

Model	ACDC val mIoU					Params(M)	FPS
	full	fog	night	rain	snow		
PIDNet-L	71.1	77.5	52.2	68.8	74.7	36.9	47
SN-FPN-MNv4	70.1	77.0	51.4	68.2	75.3	13.2	14
PIDNet-M	70.0	76.6	51.7	71.1	71.8	28.5	59
PIDNet-S	68.8	74.6	52.2	65.9	72.0	7.6	114
SN-FPN-RN18	67.7	74.8	50.1	65.2	68.2	16.7	33
SF-RN18	66.9	74.1	48.8	65.6	70.4	12.9	34
SF-lite-RN18	64.8	71.6	47.8	63.0	68.5	12.3	34
SegFormer-B0*	64.4	71.3	47.0	61.9	68.8	3.7	25*

Table 4.2: ACDC realtime model baseline. \*Model trained in PaddleSeg, latency measured in torch(timm) for fair comparison.

## Baselines (SOTA network)

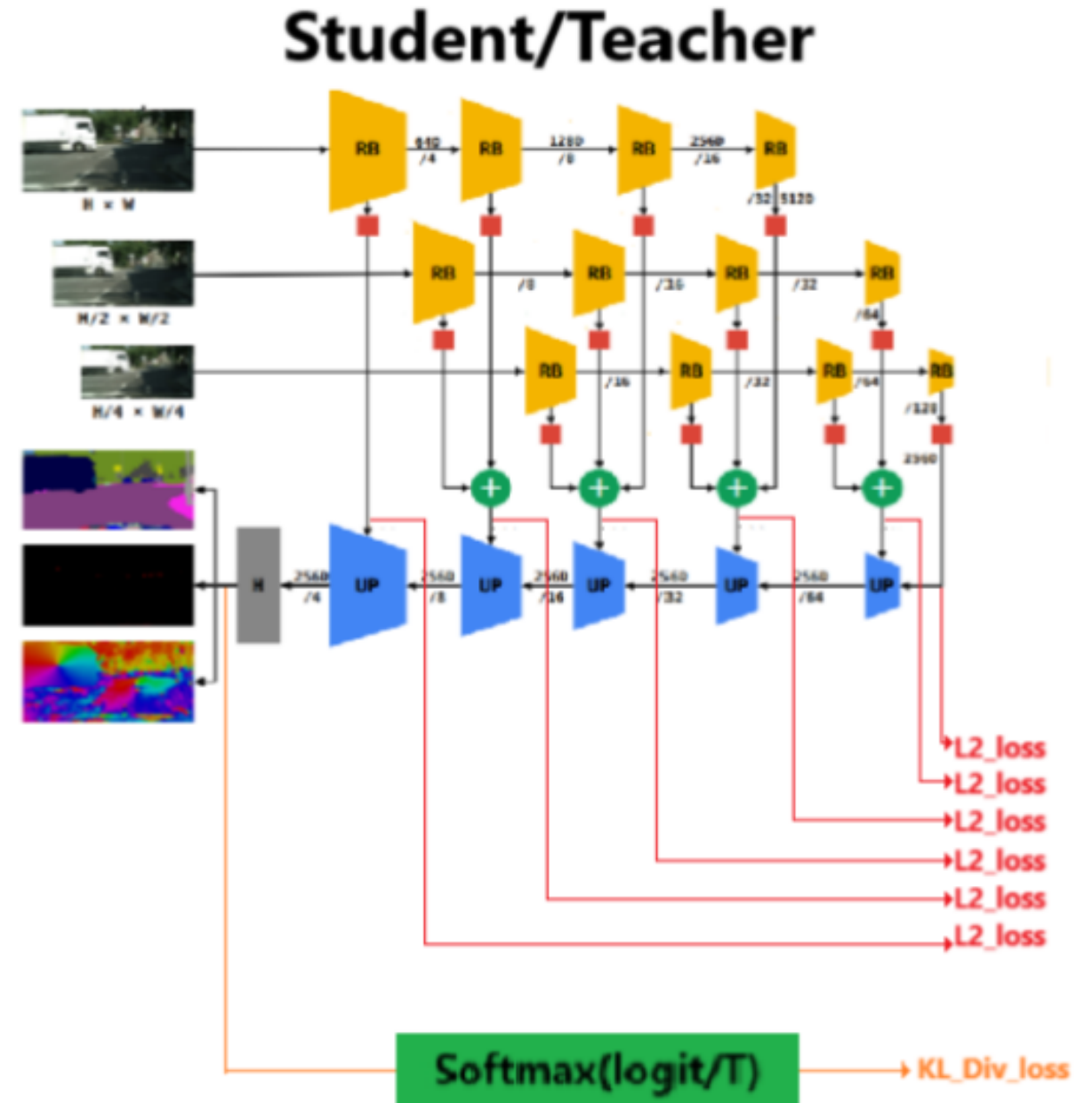
Model	ACDC val mIoU					Params(M)	FPS
	full	fog	night	rain	snow		
SWiftNet-ConvNext-L	85.0	87.7	74.0	82.1	84.0	206	4.8
HRNetv2-W48*	73.5*	74.7*	65.3*	77.7*	76.3*	66	9.5

Table 4.3: Baselines on ACDC val. \*From [\[13\]](#)

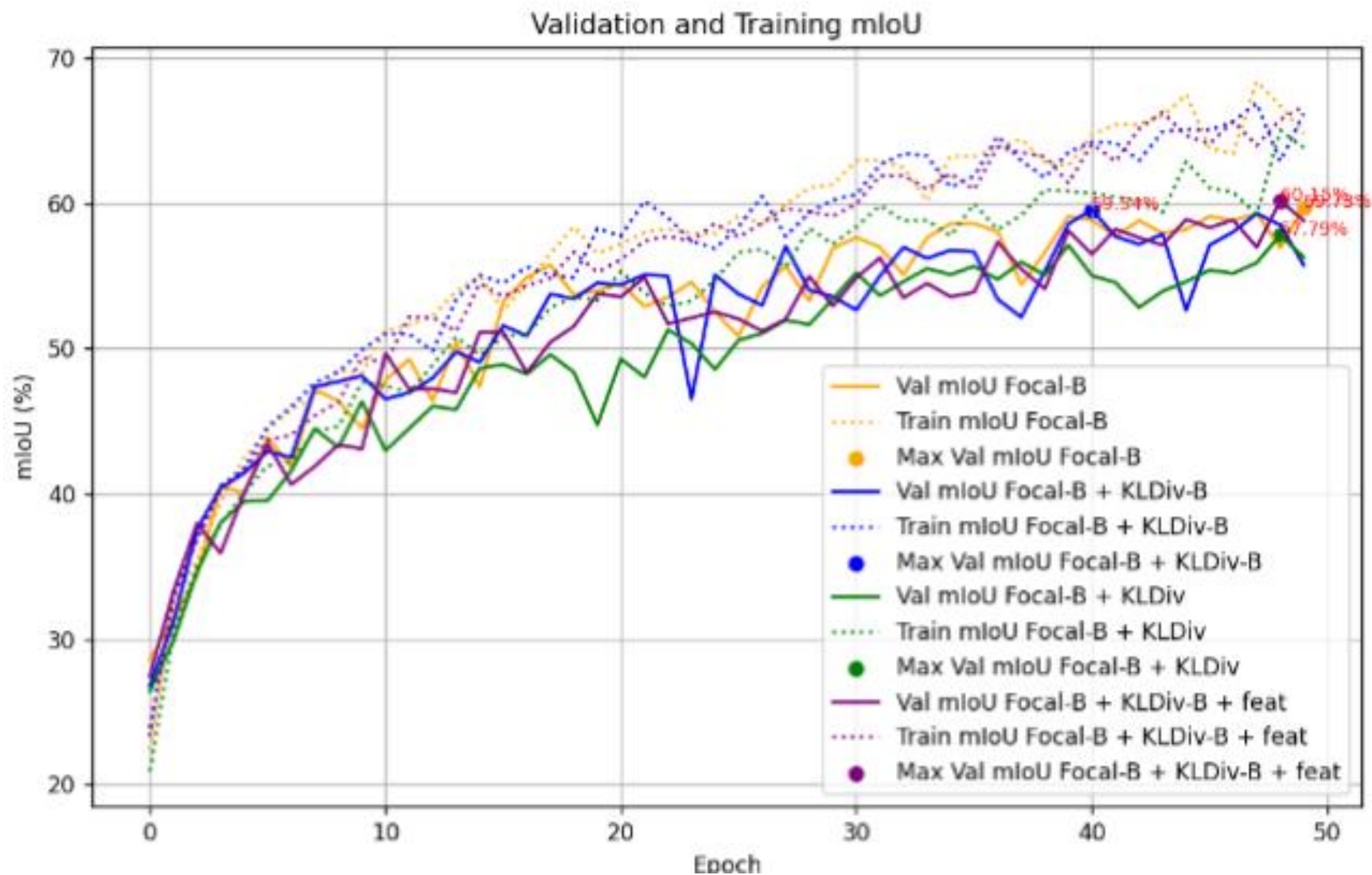


# Distillation framework for SwiftNet:

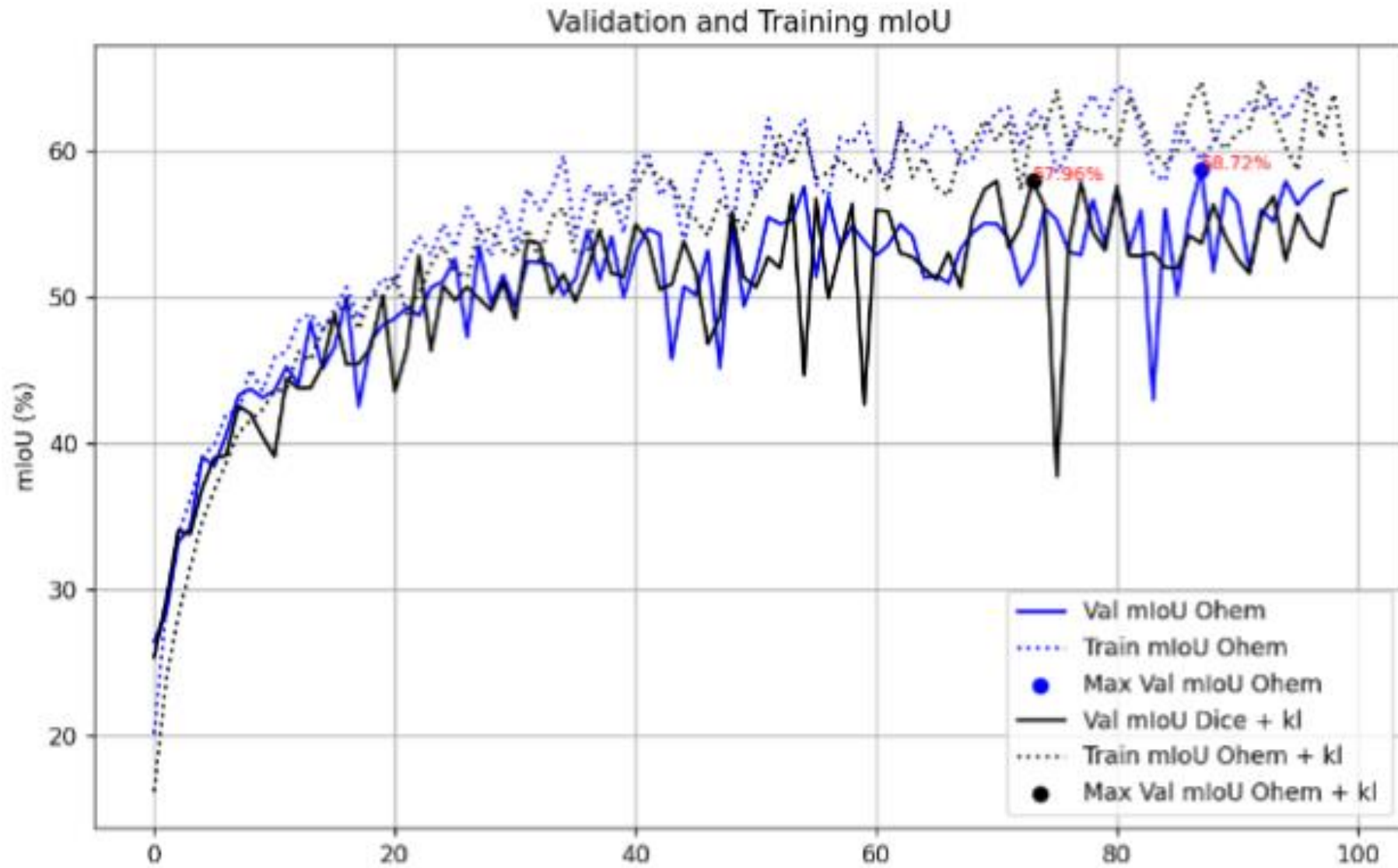
- L2 or other losses as feature loss
- Border focal loss (not in image)
- KL-Divergence on logit (perform better when teacher-student gap is low)



# Performances of logit kd (SwiftNet)



# Performances of logit kd (PIDNet)



# Loss functions for intermediate feature distillation

- **L2[1]**: match intermediate feature of teacher and student as in a regression, the objective is to make the student emulate teacher internal representation;
- **LAD[2]**: exclude the regression of magnitude (intermediate feature) between feature maps normalizing by layer;
- **CWD[3]**: channel wise distillation loss consist of applying KL-Divergence along spatial dimensions of feature maps instead of along the channel dimension, penalizing more the error in channels that focus on small (portion of) objects.

[1] FitNets: Hints for Thin Deep Nets,

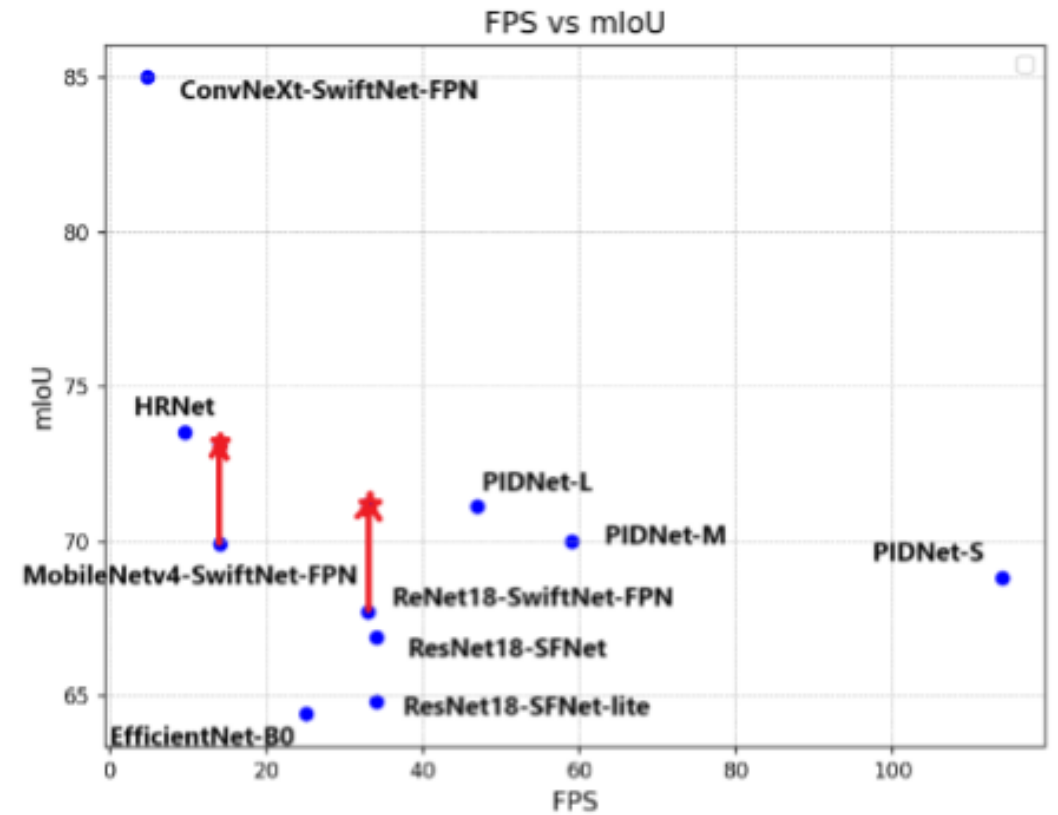
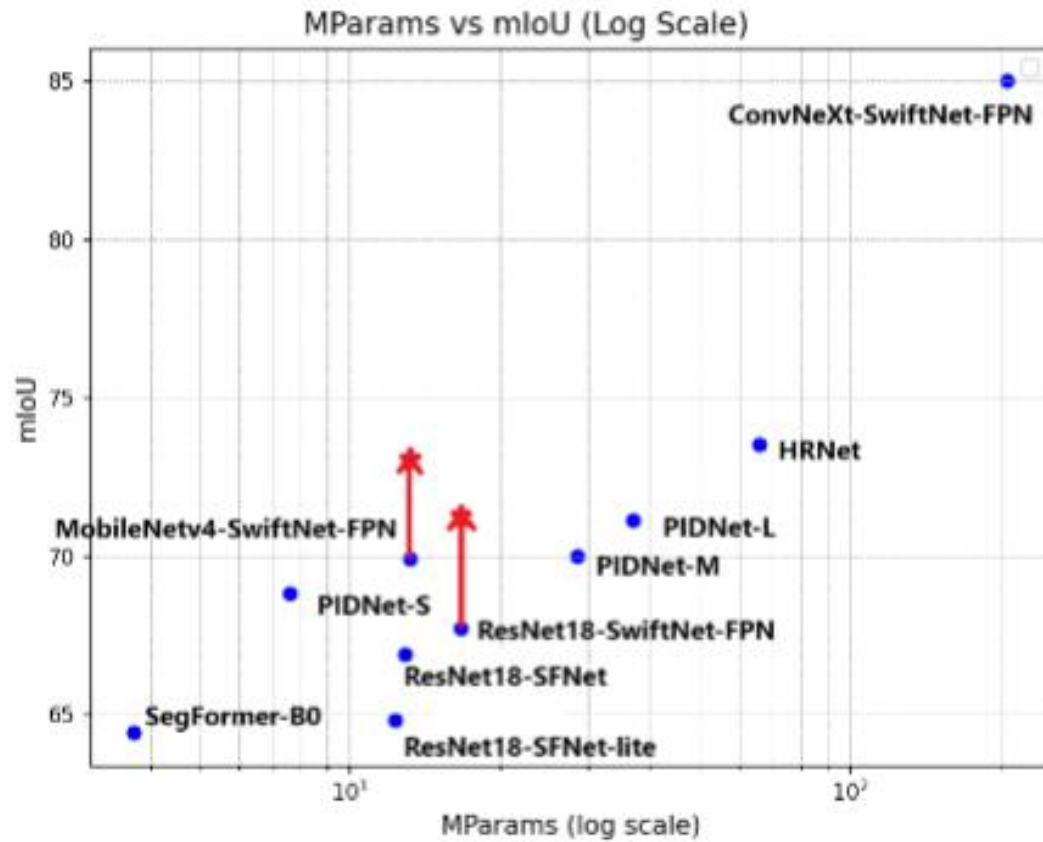
[2] Rethinking Knowledge Distillation with Raw Features for Semantic Segmentation

[3] Channel-wise Knowledge Distillation for Dense Prediction

## Intermediate feature distillation results

Backbone	Loss	ACDC val mIoU					ECE
		full	fog	night	rain	snow	
MobileNetv4	border	69.9	77.3	52.0	74.1	67.1	0.076
MobileNetv4	L2+border	72.4 (+2.5)	76.4	55.0	76.4	70.6	0.071
MobileNetv4	CWD+kl+border	72.7 (+2.7)	77.6	55.9	76.6	69.7	0.068
MobileNetv4	L2+kl+border	73.1 (+3.2)	78.1	55.2	76.3	71.6	0.070
ResNet18	border	67.7	75.7	50.5	66.3	68.7	0.059
ResNet18	L2+kl+border	68.1 (+0.4)	74.6	52.0	65.6	70.4	0.059
ResNet18	LAD+border	68.8 (+1.1)	76.5	52.7	66.6	72.5	0.065
ResNet18	L2+border	69.1 (+1.4)	75.1	52.8	67.4	71.5	0.061
ResNet18	CWD+kl+border	70.4 (+2.7)	75.7	53.7	68.2	73.4	0.059
ResNet18	CWD+border	71.2 (+3.5)	75.7	54.6	69.8	74.6	0.063

# Performances and resource usage



# Conclusions

The distillation framework used along with the change of the backbone allowed me to obtain an architecture with 1/5 of memory occupation of HRNet, with similar performances in term of mIoU (73%).

Real-time network obtained with CWD distillation reached performances comparable with PIDNet-L.

# Future works

One straightforward method to obtain higher mIoU is to perform knowledge distillation using additional datasets such as Cityscapes and BDD100K.

Another way to obtain a well-performing network can consist into distilling PIDNet, that already perform well in term of latency, but this will require more careful match between features.



Thanks for your attention!

Feel free to ask your questions

