Incorporating depth information into few-shot semantic segmentation

Yifei Zhang[1], Désiré Sidibé[2], Olivier Morel[1], Fabrice Meriaudeau[1]*

[1]ERL VIBOT CNRS 6000, ImViA, Université Bourgogne Franche Comté, 71200, Le Creusot, France
[2]Université Paris-Saclay, Univ Evry, IBISC, 91020, Evry, France
[*]Yifei.Zhang@u-bourgogne.fr

ICPR 2020, Milan, Italy

Summary

- Introduction on multimodal few-shot segmentation
- Proposed model
- Dataset
- Experimental results
- Conclusion and future work

Introduction on multimodal few-shot segmentation

Few-shot semantic segmentation presents a significant challenge for semantic scene understanding under limited supervision. Namely, this task targets at generalizing the segmentation ability of the model to new categories given a few samples. In order to obtain complete scene understanding, we extend the RGB-centric methods to take advantage of complementary depth information.

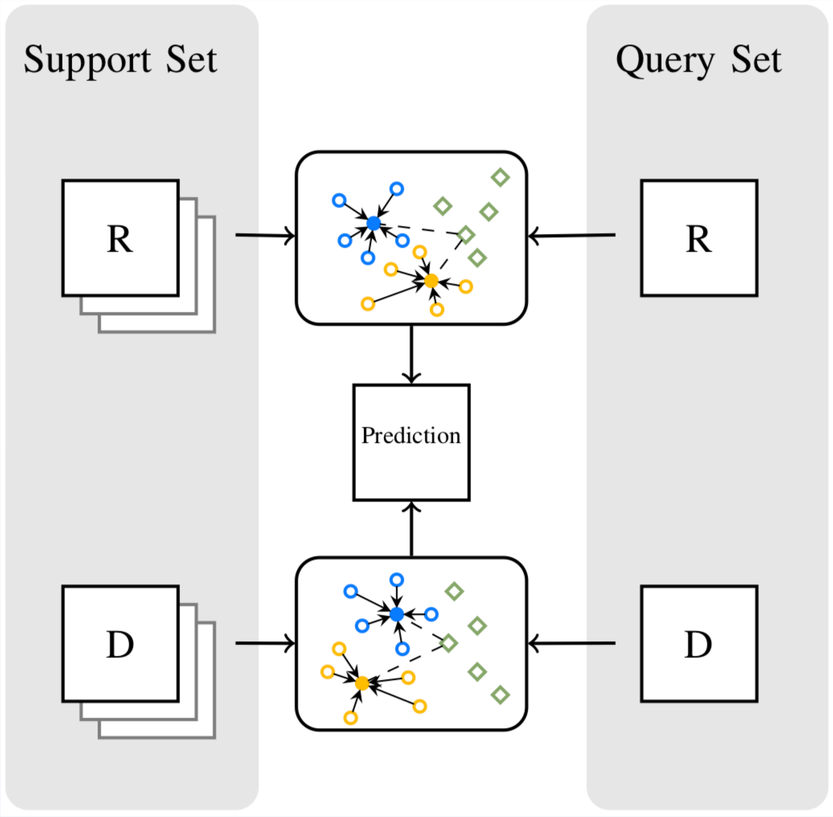

Figure 1: An overview of the proposed method (RDNet).

Summary

- Introduction on multimodal few-shot segmentation
- Proposed model
- Dataset
- Experimental results
- Conclusion and future work

Proposed model

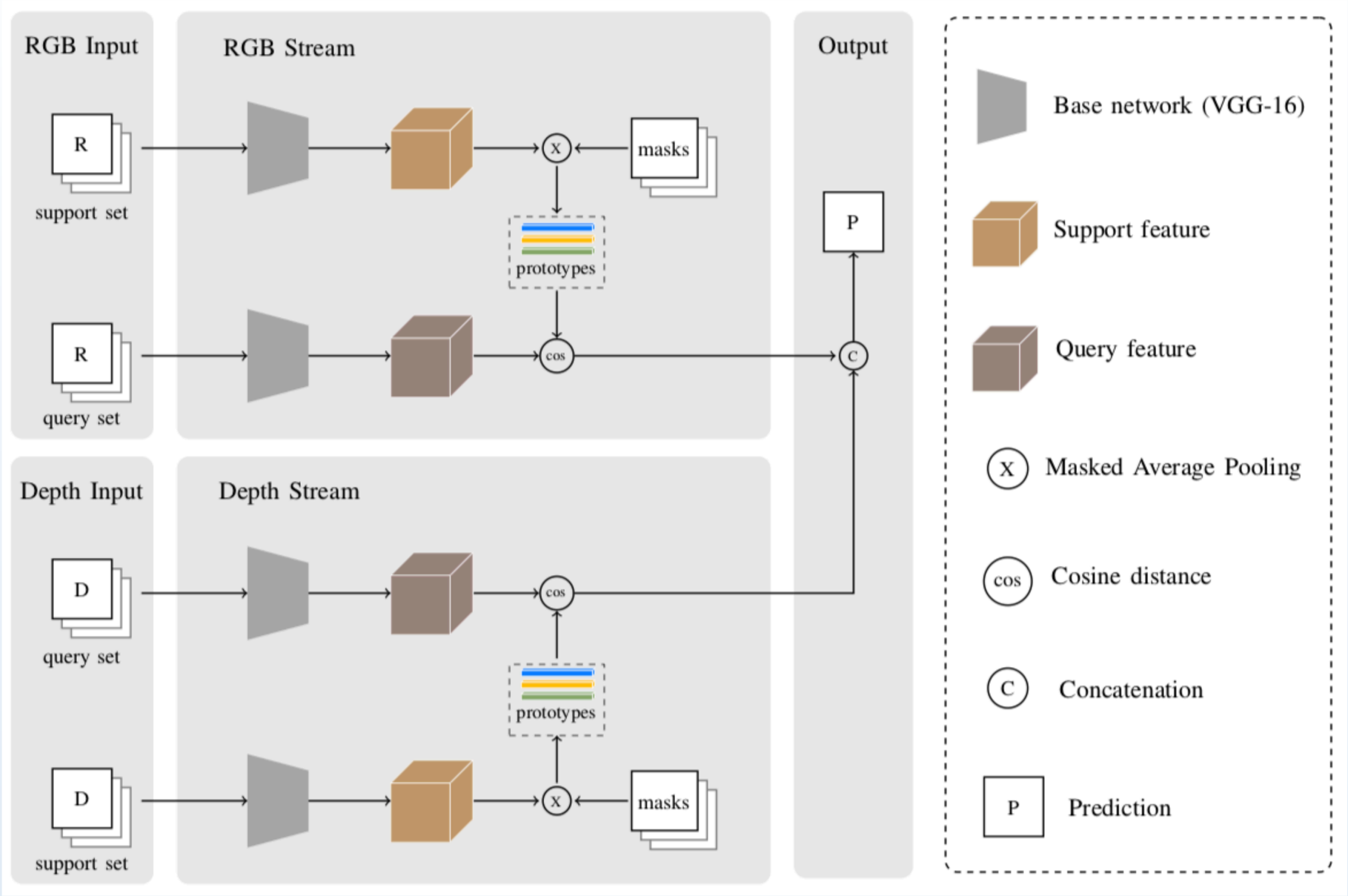

Figure 2: Illustration of the proposed method (RDNet)

Summary

- Introduction on multimodal few-shot segmentation
- Proposed model
- Dataset
- Experimental results
- Conclusion and future work

Dataset

Dataset	Test classes
Cityscapes-3^0	road, sidewalk, bus
Cityscapes-3^1	vegetation, terrain, sky
Cityscapes-3^2	human, car, building

Figure 3: Training and evaluation on Cityscapes-3dataset using 3-fold cross-validation

Figure 4: Example image from the PASCAL VOC dataset.

Figure 5: Example image from the Cityscapes dataset.

Summary

- Introduction on multimodal few-shot segmentation
- Proposed model
- Dataset
- Experimental results
- Conclusion and future work

Experimental results

Methods	Modality	1-way 1-shot				1-way 2-shot			
		Cityscapes-3^0	Cityscapes-3^1	Cityscapes-3^2	Mean	Cityscapes-3^0	Cityscapes-3^1	Cityscapes-3^2	Mean
PANet	RGB	35.2	19.7	32.1	29.0	37.2	23.2	36.7	32.4
RDNet-R		35.7	22.3	32.6	30.2	36.7	24.1	37.5	32.8
PANet	Depth	32.6	14.5	19.3	22.1	34.2	15.8	22.5	24.2
RDNet-D		35.1	15.8	21.0	24.0	33.7	17.3	25.3	25.4
RDNet-concat	RGB-D	33.8	15.7	20.7	23.4	34.3	17.9	26.9	26.4
RDNet (ours)		**36.8**	**23.5**	**33.3**	**31.2**	**37.3**	**26.1**	**37.6**	**33.7**

Table 1: Results of 1-way 1-shot and 1-way 2-shot semantic segmentation on Cityscapes-3^i using mean-IoU (%) metric.

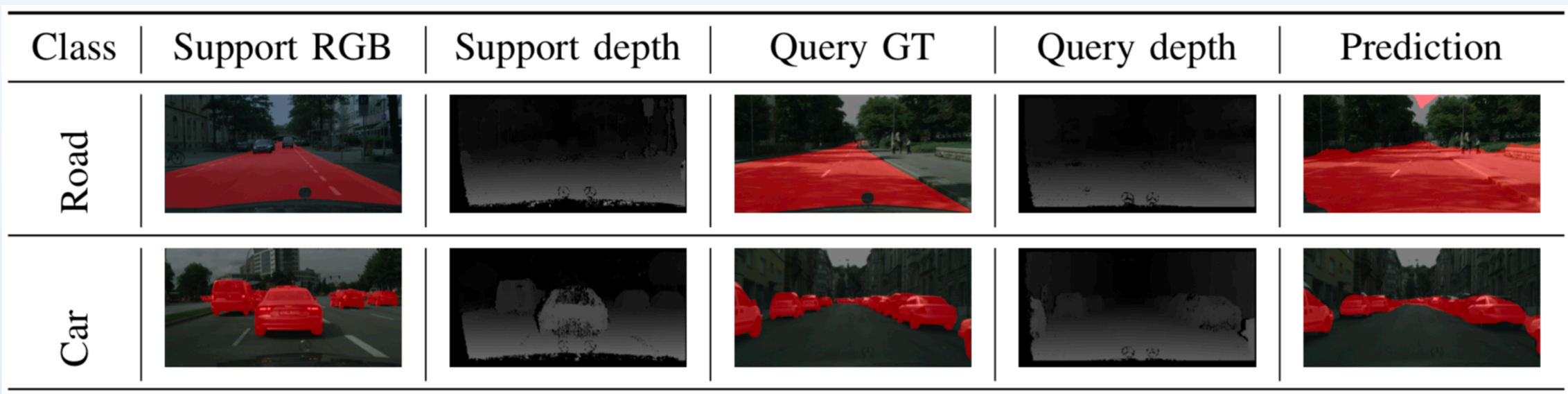

Table 2: Qualitative results of our method for 1-way 1-shot semantic segmentation on Cityscapes-3^i.

Experimental results

Class	RDNet	RDNet-R	RDNet-D
Mean	**31.2**	30.2	24.0
Road	83.0	80.9	**84.4**
Sidewalk	**17.8**	15.7	15.7
Bus	9.5	**10.6**	5.3
Vegetation	**43.1**	40.2	26.9
Terrain	8.3	**10.1**	6.8
Sky	**19.1**	16.7	13.7
Human	**47.8**	46.6	36.9
Car	**12.1**	12.1	5.0
Building	**39.9**	39.2	21.1

Table 3: Per-class mean-IoU (%) comparison of ablation studies for 1-way 1-shot semantic segmentation.

Mehtods	Modality	binary IoU	Runtime
PANet	RGB	55.0	71ms
RDNet-R		56.5	65ms
RDNet-concat	RGB-D	51.9	67ms
RDNet (ours)		57.9	135ms

Table 4: Results of 1-way 1-shot semantic segmentation using binary-IoU and the runtime.

Experimental results

Visualization using T-SNE:

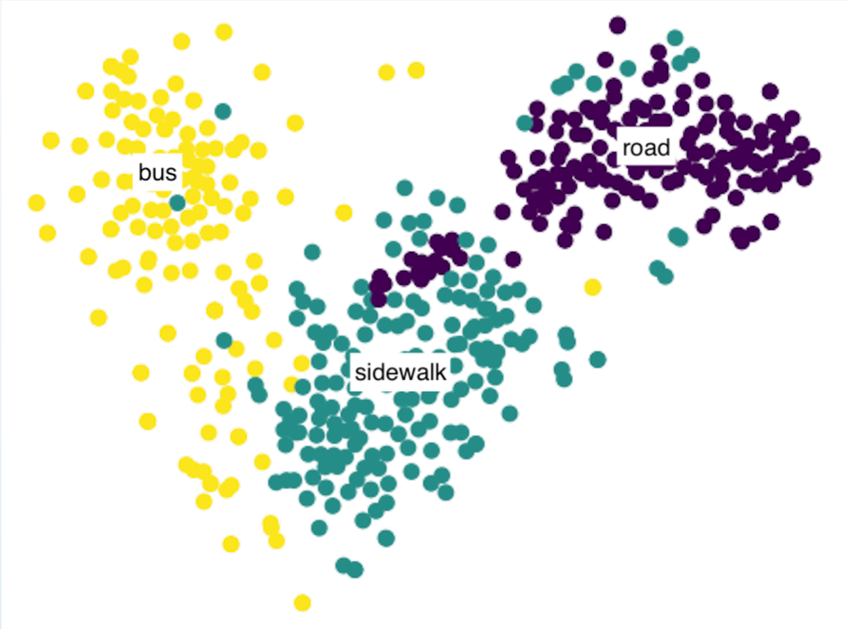

Figure 6: RGB embeddings in Cityscapes-3^0.

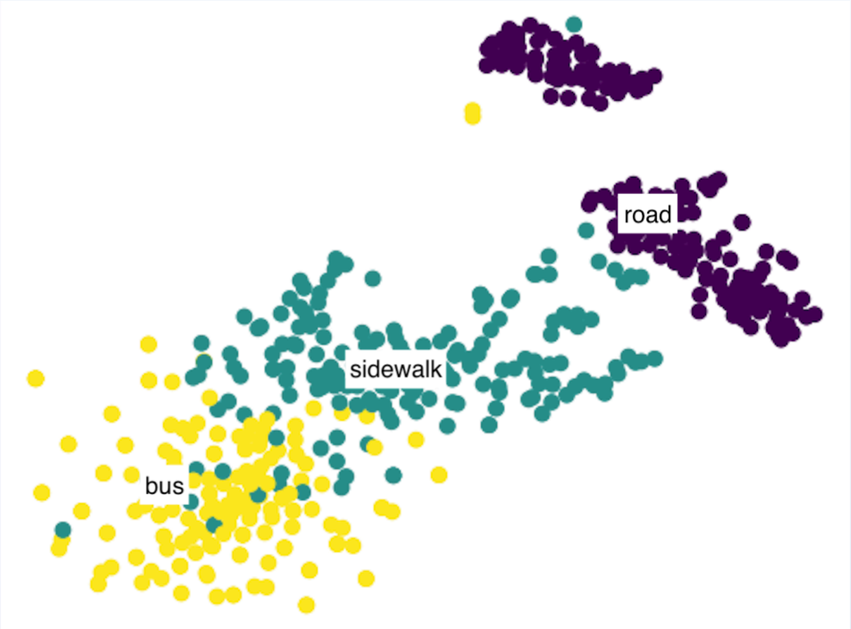

Figure 7: Depth embeddings in Cityscapes-3^0.

Summary

- Introduction on multimodal few-shot segmentation
- Proposed model
- Dataset
- Experimental results
- Conclusion and future work

Conclusion and future work

Conclusion:

- Comprehensive experiments and ablation studies on Cityscapes-3^idataset demonstrate the improved generalizability and discriminating ability of our method.
- The proposed method is simple yet effective, and explore the positive use of depth information in few-shot segmentation tasks.

Future work:

- The integration of different multimodal information in few-shot learning tasks

Thank you!