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

INTRODUCTION

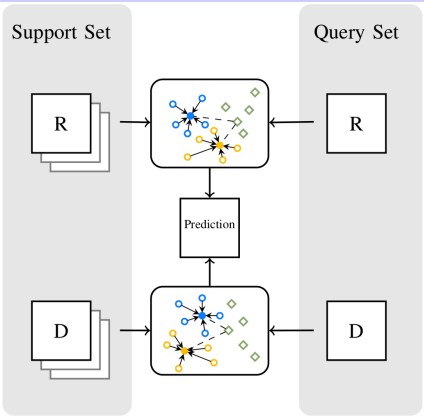
Few-shot segmentation presents a significant challenge for semantic scene understanding under limited supervision. Namely, this task targets at generalizing the segmentation ability of the model to new categories given a few samples. In order to obtain complete scene understanding, we extend the RGB-centric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

Contribution:

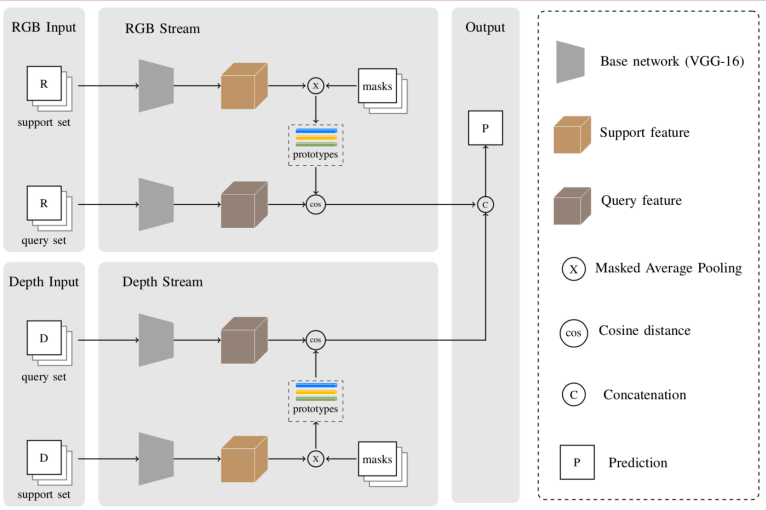
(1) We propose a metric learning-based deep neural network for few-shot semantic segmentation, which processes RGB-D data in two streams.

(2) We define a new few-shot segmentation benchmark on the Cityscapes dataset, named Cityscapes-3^i.

(3) Extensive experiments and ablation studies demonstrate the effectiveness of the proposed RDNet, as well as the positive effects of geometric information in limited supervisory scene understanding.



MODEL



The proposed RDNet architecture includes two mirrored streams: an RGB stream and a depth stream. Each stream processes the corresponding input data, including a support set and a query set. The prototypes of support images are obtained by masked average pooling. Then the semantic guidance is performed on the query feature by computing the relative cosine distance. The results from these two streams are combined at the late stage.

DATASET&SETTING

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

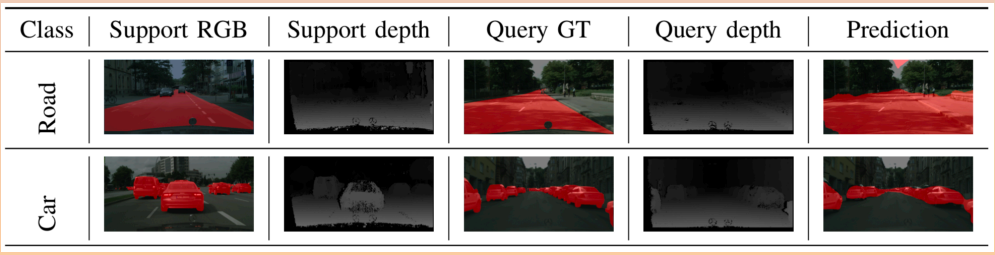
Training and evaluation on Cityscapes-3^i dataset using 3-fold cross-validation

- $D_{train}=(x_i^R, x_i^D, y(l)_i)^N_{i=1}$
- $D_s=(x_j^R, x_j^D, y(l)_i)^M_{i=1}$
- $D_q=(x_j^R, x_j^D)^n_{i=1}$

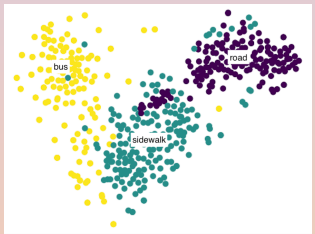
EXPERIMENTAL RESULTS

Methods	Modality	1-way 1-shot			
		Cityscapes-3^0	Cityscapes-3^1	Cityscapes-3^2	Mean
PANet	RGB	35.2	19.7	32.1	29.0
RDNet-R		35.7	22.3	32.6	30.2
PANet	Depth	32.6	14.5	19.3	22.1
RDNet-D		35.1	15.8	21.0	24.0
RDNet-concat	RGB-D	33.8	15.7	20.7	23.4
RDNet (ours)		**36.8**	**23.5**	**33.3**	**31.2**

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



VISUALIZATION USING T-SNE



RGB embeddings in Cityscapes-3^0

Depth embeddings in Cityscapes-3^0