PANet：Few-Shot Image Semantic Segmentation with Prototype Alignment

2019,ICCV

问题背景

传统语义分割模型如FCN，SegNet，DeepLab，PSPNet等，训练模型都需要大量带标注的图像，成本高并且最后训练出的模型对未知类别的推广能力也很差。因此需要对小样本语义分割方法进行研究。

作者指出当前小样本语义分割存在的问题：

① 没有区分特征提取和分割的过程，作者觉得分割模型表示和支持集的语义特征混合在一起是有问题的。

② 仅仅把支持集的注释用来生成掩码。（应该是指没有很好利用支持集的信息）

问题背景

① 针对第一个问题，作者将这两个部分分离为原型提取和非参数度量学习。
② 针对第二个问题，通过反向执行小样本语义分割实现一种新的原型对齐正则化，即将查询图像(query iamge)和预测掩码(predicted mask)作为一个新的支持集用来分割之前的支持图像。使得模型在支持和查询之间生成更好的原型，获得更好的泛化能力。

问题设置

支持集S：$\{I_{c,k}, M_{c,k}\}$，用来表示一个C-way K-shot的分割问题（K-shot表示对一类需要有几张图来训练，C-way表示这个框架一次要识别C个类别）。

查询集Q：$\{iamge, mask\}$，一个查询图像和掩码对，其中，查询图像中包含要识别的C个类别。

PANet

网络框架

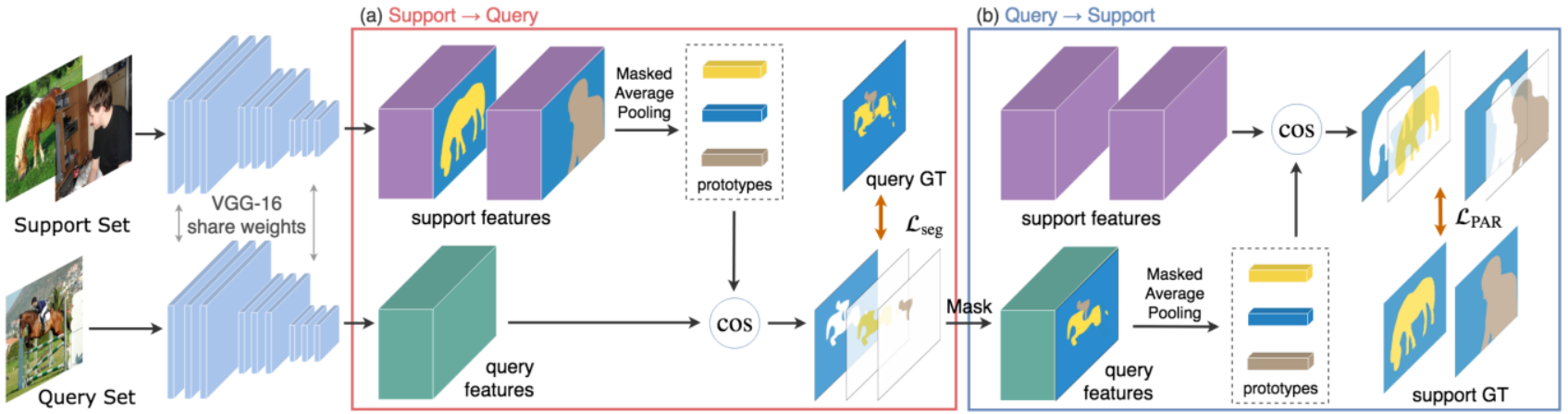


Figure 2: Illustration of the pipeline of our method in a 2-way 1-shot example. In block (a), PANet performs a support-to-query few-shot segmentation. The support and query images are embedded into deep features. Then the prototypes are obtained by masked average pooling. The query image is segmented via computing the cosine distance (cos in the figure) between each prototype and query features at each spatial location. Loss $\mathcal{L}_{\text{seg}}$ is computed between the segmentation result and the ground truth mask. In block (b), the proposed PAR aligns the prototypes of support and query by performing a query-to-support few-shot segmentation and calculating loss $\mathcal{L}_{\text{PAR}}$. GT denotes the ground truth segmentation masks.

PANet

网络框架

①首先利用同一个backbone(VGG-16)对支持集和查询集的图像进行特征提取，然后对支持集产生的特征图进行一个掩码操作。

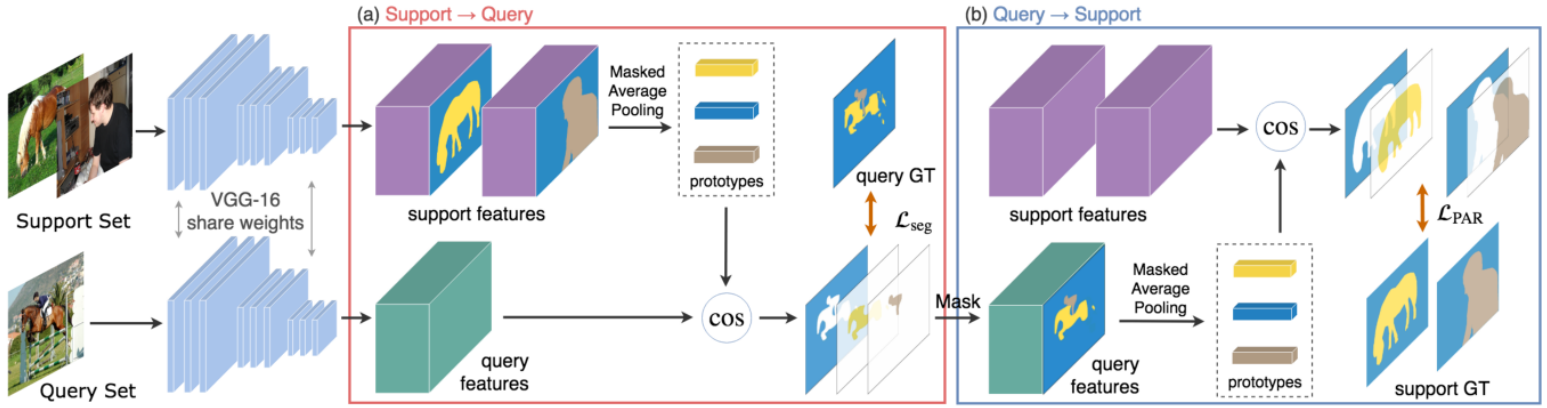


Figure 2: Illustration of the pipeline of our method in a 2-way 1-shot example. In block (a), PANet performs a support-to-query few-shot segmentation. The support and query images are embedded into deep features. Then the prototypes are obtained by masked average pooling. The query image is segmented via computing the cosine distance (cos in the figure) between each prototype and query features at each spatial location. Loss $\mathcal{L}_{seg}$ is computed between the segmentation result and the ground truth mask. In block (b), the proposed PAR aligns the prototypes of support and query by performing a query-to-support few-shot segmentation and calculating loss $\mathcal{L}_{PAR}$. GT denotes the ground truth segmentation masks.

PANet

网络框架

②然后再对支持集的特征图通过平均池化的操作计算得到每个类别所对应的原型向量。计算的公式如下:

$$p_c = \frac{1}{K}\sum_k \frac{\sum_{x,y} F_{c,k}^{(x,y)} \mathbb{1}[M_{c,k}^{(x,y)} = c]}{\sum_{x,y} \mathbb{1}[M_{c,k}^{(x,y)} = c]},$$

$$p_{\text{bg}} = \frac{1}{CK}\sum_{c,k} \frac{\sum_{x,y} F_{c,k}^{(x,y)} \mathbb{1}[M_{c,k}^{(x,y)} \notin \mathcal{C}_i]}{\sum_{x,y} \mathbb{1}[M_{c,k}^{(x,y)} \notin \mathcal{C}_i]}.$$

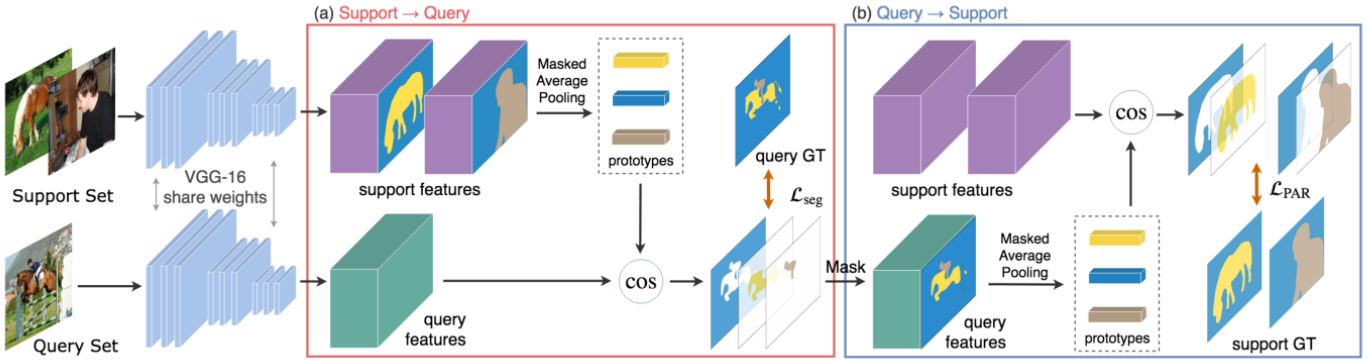


Figure 2: Illustration of the pipeline of our method in a 2-way 1-shot example. In block (a), PANet performs a support-to-query few-shot segmentation. The support and query images are embedded into deep features. Then the prototypes are obtained by masked average pooling. The query image is segmented via computing the cosine distance (cos in the figure) between each prototype and query features at each spatial location. Loss $\mathcal{L}_{\text{seg}}$ is computed between the segmentation result and the ground truth mask. In block (b), the proposed PAR aligns the prototypes of support and query by performing a query-to-support few-shot segmentation and calculating loss $\mathcal{L}_{\text{PAR}}$. GT denotes the ground truth segmentation masks.

PANet

网络框架

③用之前从支持集中生成的原型向量对查询集的特征图进行一个预测，方法是先计算查询集每个位置的特征向与各个类别的原型向量之间的距离，再用一个softmax函数转换为概率值。

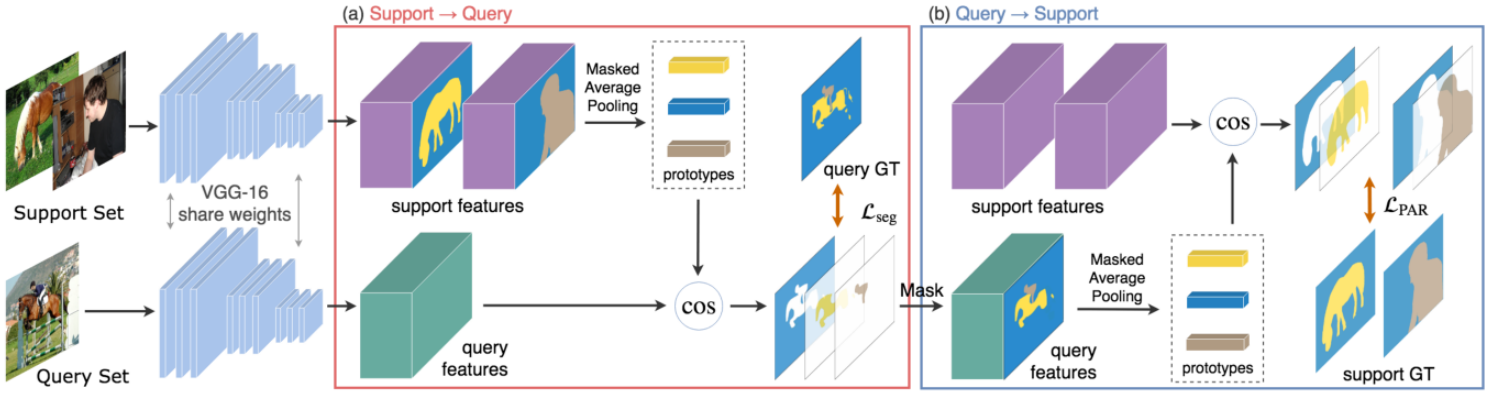


Figure 2: Illustration of the pipeline of our method in a 2-way 1-shot example. In block (a), PANet performs a support-to-query few-shot segmentation. The support and query images are embedded into deep features. Then the prototypes are obtained by masked average pooling. The query image is segmented via computing the cosine distance (cos in the figure) between each prototype and query features at each spatial location. Loss $\mathcal{L}_{seg}$ is computed between the segmentation result and the ground truth mask. In block (b), the proposed PAR aligns the prototypes of support and query by performing a query-to-support few-shot segmentation and calculating loss $\mathcal{L}_{PAR}$. GT denotes the ground truth segmentation masks.

$$\tilde{M}_{q;j}^{(x,y)} = \frac{\exp(-\alpha d(F_q^{(x,y)}, p_j))}{\sum_{p_j \in \mathcal{P}} \exp(-\alpha d(F_q^{(x,y)}, p_j))}.$$

$$\hat{M}_q^{(x,y)} = \arg\max_j \tilde{M}_{q;j}^{(x,y)}.$$

PANet

网络框架

④PAR方法：上述三步操作已经完成了对查询图像的一个预测。但是作者提出**在训练过程中使用**查询图像的预测掩码和特征图当作支持集，反过来对原本的支持集(Support Set)进行一个预测。预测的方法就和前三步操作是一样的。这样能够从支持集中学到更多有用的信息。

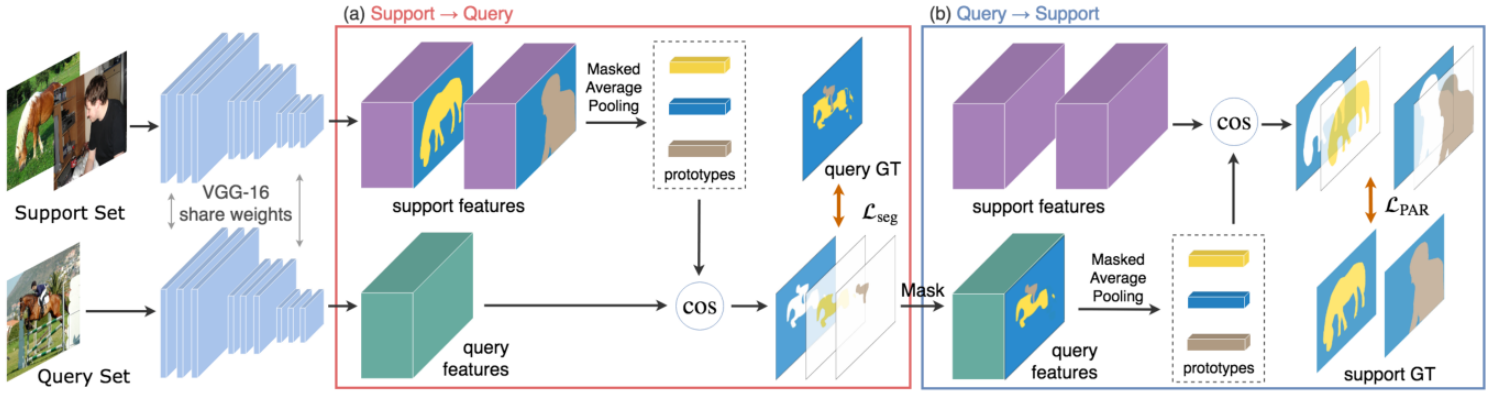


Figure 2: Illustration of the pipeline of our method in a 2-way 1-shot example. In block (a), PANet performs a support-to-query few-shot segmentation. The support and query images are embedded into deep features. Then the prototypes are obtained by masked average pooling. The query image is segmented via computing the cosine distance (cos in the figure) between each prototype and query features at each spatial location. Loss $\mathcal{L}_{seg}$ is computed between the segmentation result and the ground truth mask. In block (b), the proposed PAR aligns the prototypes of support and query by performing a query-to-support few-shot segmentation and calculating loss $\mathcal{L}_{PAR}$. GT denotes the ground truth segmentation masks.

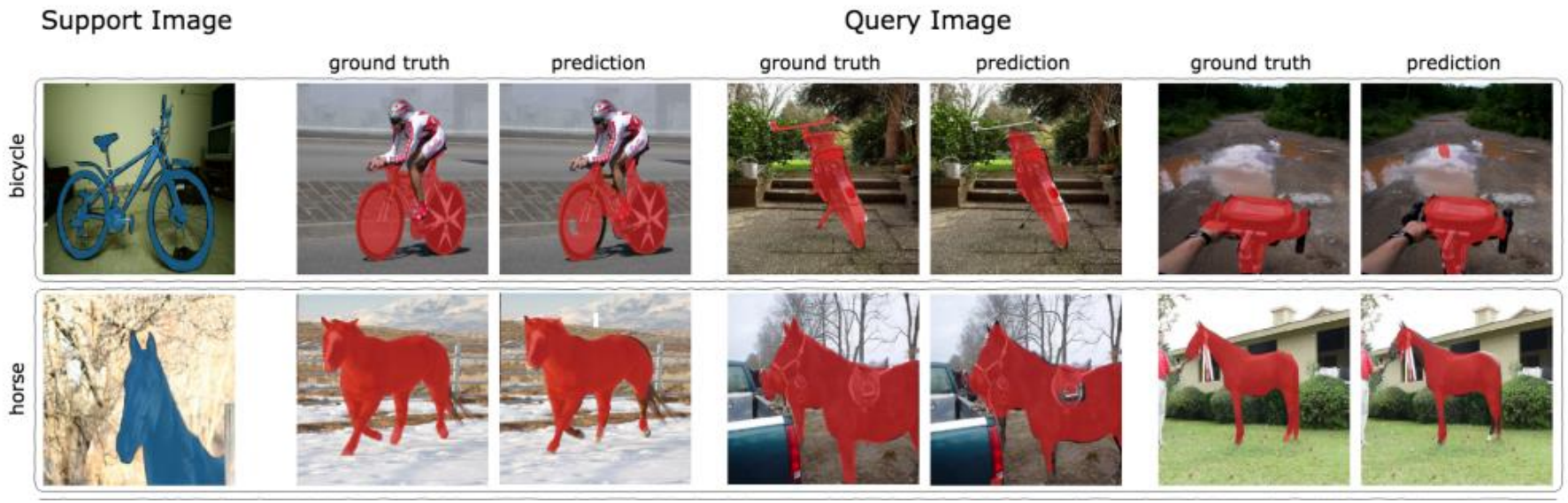
测试结果

Method	1-shot					5-shot					Δ	#Params
	split-1	split-2	split-3	split-4	Mean	split-1	split-2	split-3	split-4	Mean	Mean	
OSLSM [21]	33.6	55.3	40.9	33.5	40.8	35.9	58.1	42.7	39.1	43.9	3.1	272.6M
co-FCN [16]†	36.7	50.6	44.9	32.4	41.1	37.5	50.0	44.1	33.9	41.4	0.3	34.2M
SG-One [28]	40.2	**58.4**	48.4	38.4	46.3	41.9	58.6	48.6	39.4	47.1	0.8	19.0M
PANet-init	30.8	40.7	38.3	31.4	35.3	41.6	52.7	51.6	40.8	46.7	**11.4**	14.7M
PANet	**42.3**	58.0	**51.1**	**41.2**	**48.1**	**51.8**	**64.6**	**59.8**	**46.5**	**55.7**	7.6	**14.7M**

Table 1: Results of 1-way 1-shot and 1-way 5-shot segmentation on PASCAL-5^i dataset using mean-IoU metric. Δ denotes the difference between 1-shot and 5-shot. †: The results of co-FCN in mean-IoU metric are reported by [28].

没有使用解码器模块和后处理技术来细化结果，但已经比当时的State-of the-art高出了7.6个点。

测试结果


Support Image
Query Image
ground truth
prediction
ground truth
prediction
ground truth
prediction
bicycle
horse

测试结果

Figure 3: Qualitative results of our model in 1-way 1-shot segmentation on PASCAL-5^i (row 1 and 2) and MS COCO (row 3 and 4).

一些失败的结果：

① 模型tends to give segmentation results with unnatural patches，作者觉得是因为模型在每一个位置是独立预测的，所以较大的几乎占满一整个图片的东西就会识别的零零散散。

②模型不能很好的区分一些具有相似原型的物体，例如tables和chairs。

测试结果

消融实验:
使用 or 不使用 PAR

Method	1-shot	5-shot
PANet w/o PAR	47.2	54.9
PANet	48.1	55.7

Table 5: Evaluation results of our PANet trained with and without PAR on PASCAL-5^i in mean-IoU metric.

测试结果

在弱标注的数据集下也获得了较好的结果。

Annotations	1-shot	5-shot
Dense	48.1	55.7
Scribble	44.8	54.6
Bounding box	45.1	52.8

Table 6: Results of using different types of annotations in mean-IoU metric.

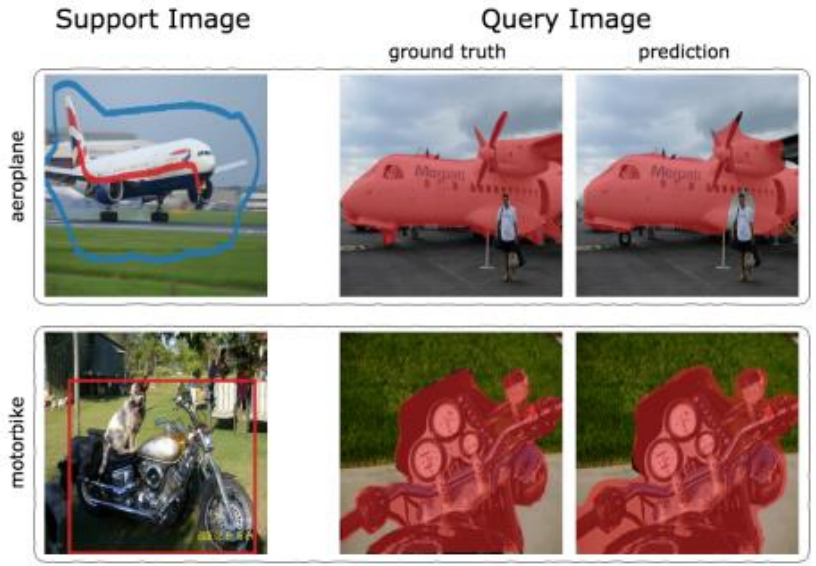


Figure 6: Qualitative results of our model on 1-way 1-shot segmentation using scribble and bounding box annotations. The scribbles are dilated for better visualization.

总结

优点：

① 采用度量学习的方法，没有引入额外的可学习参数，因此不容易过拟合。
② 原型嵌入和预测是在计算的特征图上执行的，因此分割不需要额外通过网络。
③ 设计的PAR非常有趣，用查询集的结果反过来去预测支持集的结果，二者相互印证。

局限：

① 模型不能很好的区分一些具有相似原型的物体，例如tables和chairs。
② 模型tends to give segmentation results with unnatural patches，作者觉得是因为模型在每一个位置是独立预测的，所以较大的几乎占满一整个图片的东西就会识别的零零散散。

Q & A