# Learning Non-target Knowledge for Few-shot Semantic Segmentation

CVPR,2022

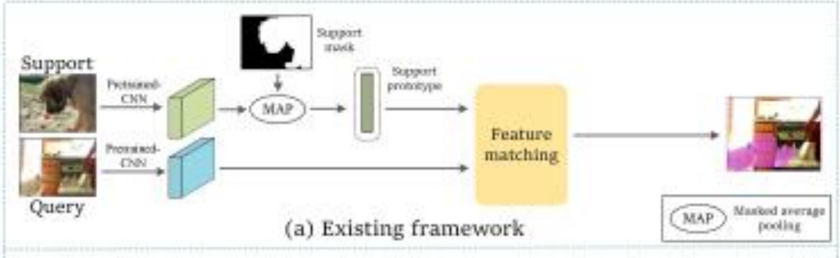
学习小样本语义分割中的非目标知识

# Motivation

现有的小样本语义分割只关注目标对象的信息挖掘，导致模型往往难以学习分辨模糊区域，背景(BG)和干扰对象(DOs).

提出了一个新的框架，用于挖掘和消除查询图像中的BG和DO区域。进而得到更好的分割结果。

Figure 1. Previous methods often show false positive predictions in non-target regions. Pixels in red indicate the target objects, while pixels in green mean false positive predictions.

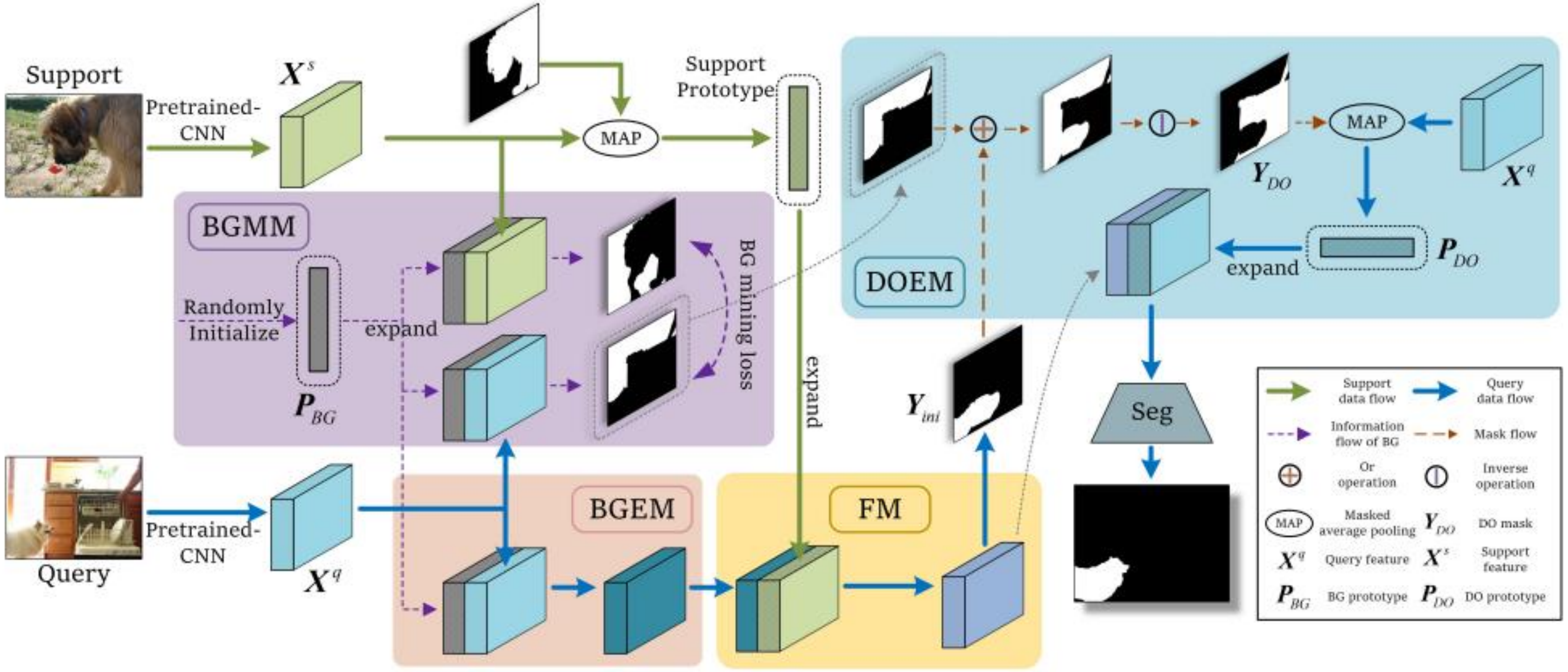



(a) Existing framework

# Works

- 提出一个NTRE网络来解决上述问题并消除 背景 和 干扰项
- 提出一个BG挖掘模块 BGMM 专门设计了损失函数
- 提出BG消除模块和DOs消除模块
- 还提出了一种原型对比学习的方法

# Framework



Support
Pretrained-CNN
$X^s$
MAP
Support Prototype
BGMM
Randomly Initialize
expand
$P_{BG}$
BG mining loss
expand
DOEM
$Y_{DO}$
MAP
$X^q$
$P_{DO}$
expand
$Y_{ini}$
Seg
Query
Pretrained-CNN
$X^q$
BGEM
FM
Support data flow
Query data flow
Information flow of BG
Mask flow
Or operation
Inverse operation
Masked average pooling
DO mask
Query feature
Support feature
BG prototype
DO prototype

# Framework

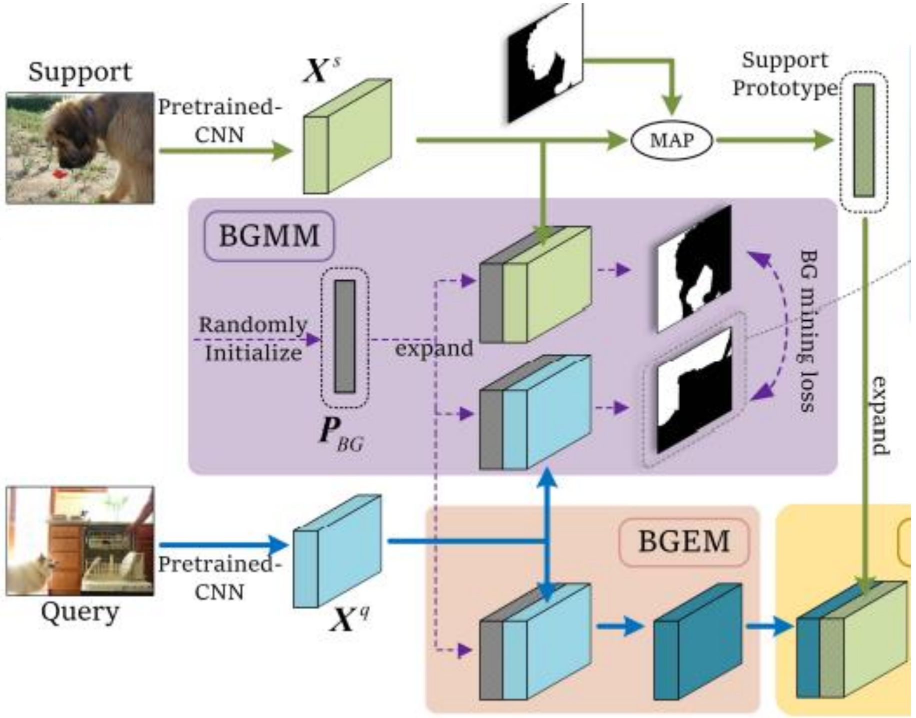
BGMM:

1.随机初始化一个 1×1×C的背景原型
2.然后将其分别与 $X_q$和 $X_s$连接在一起 再通过两个3×3的卷积 得到支持集和查询集的BG预测
3.两个卷积块共享相同的权重

$$\boldsymbol{y}_{BG}^{q} = \mathcal{F}_{3\times3}(\boldsymbol{X}^{q} \oplus \hat{\boldsymbol{P}}_{BG}), \tag{1}$$

$$\boldsymbol{y}_{BG}^{s} = \mathcal{F}_{3\times3}(\boldsymbol{X}^{s} \oplus \hat{\boldsymbol{P}}_{BG}), \tag{2}$$

$$\begin{aligned} L_{BG} = &- \frac{1}{N}\sum_{i} log(1 - \boldsymbol{y}_{BG}^{q/s}(i))\boldsymbol{M}^{q/s}(i) \\ &- \alpha\frac{1}{Z}\sum_{j} log(\boldsymbol{y}_{BG}^{q/s}(j)), \end{aligned} \tag{3}$$

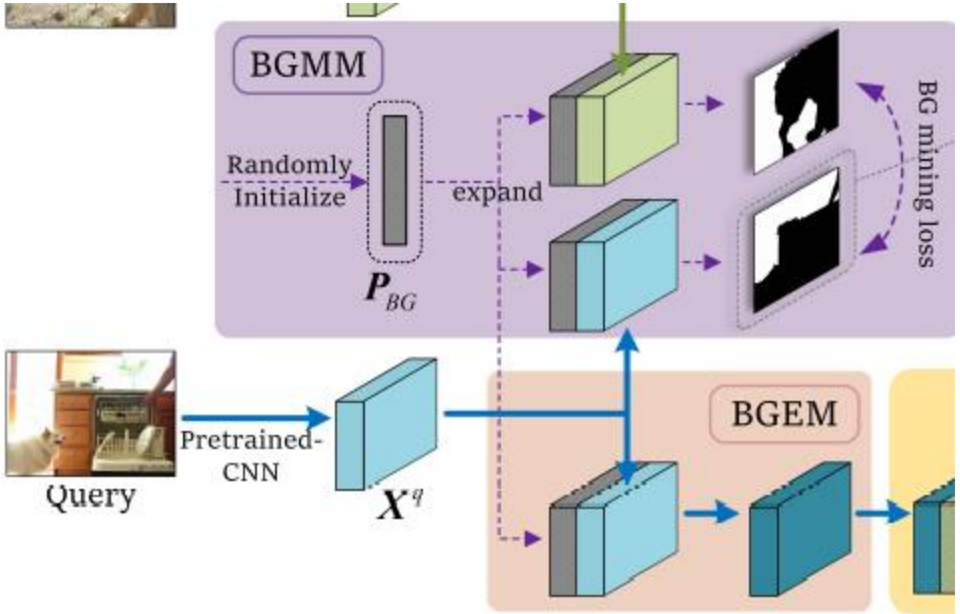


# Framework

BGEM:

1.将背景原型 与 查询特征$X_q$进行连接
2.使用一个1×1的卷积层 排除查询特征中的BG信息

$$\boldsymbol{X}_{BG}^{q} = \mathcal{F}_{1\times1}(\boldsymbol{X}^{q} \oplus \hat{\boldsymbol{P}}_{BG}), \qquad (4)$$



# Framework

支持特征匹配:

1.先在支持集特征图上进行MAP操作 得到支持原型 $P_s \in 1 \times 1 \times C$

2.将其扩展为 H×W×C 并与BG过滤的查询特征连接

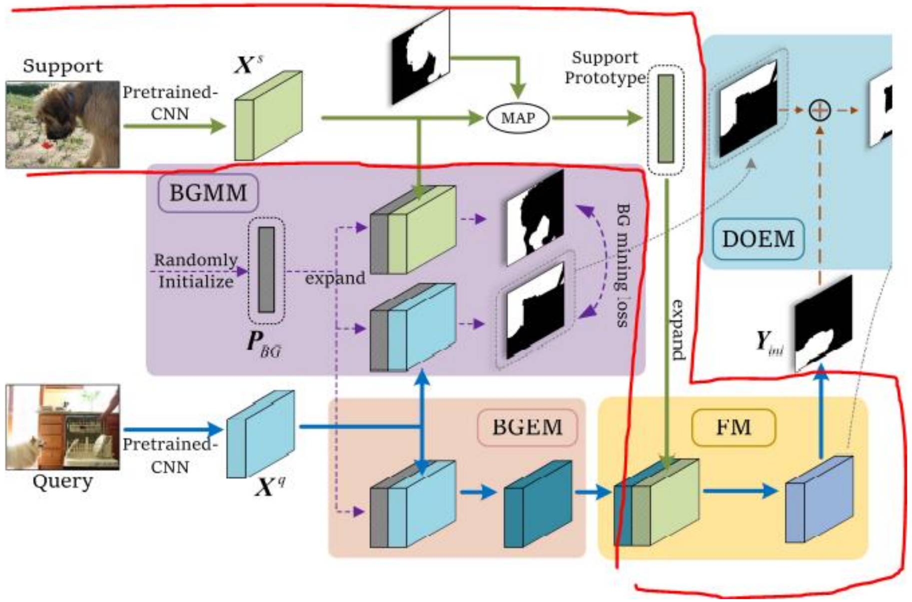
3.还引入先验置信图 $C_p$ ∈ H×W×1

4.再通过一个1×1的卷积  得到激活的查询特征

5.再通过两个3×3的卷积块 得到初步预测结果

$$\boldsymbol{X}_{act}^{q} = \mathcal{F}_{1\times1}(\boldsymbol{X}_{BG}^{q} \oplus \hat{\boldsymbol{P}}^{s} \oplus \boldsymbol{C}_{p}), \qquad (5)$$

$$\boldsymbol{y}_{ini}^{q} = \mathcal{F}_{3\times3}(\boldsymbol{X}_{act}^{q}), \qquad (6)$$



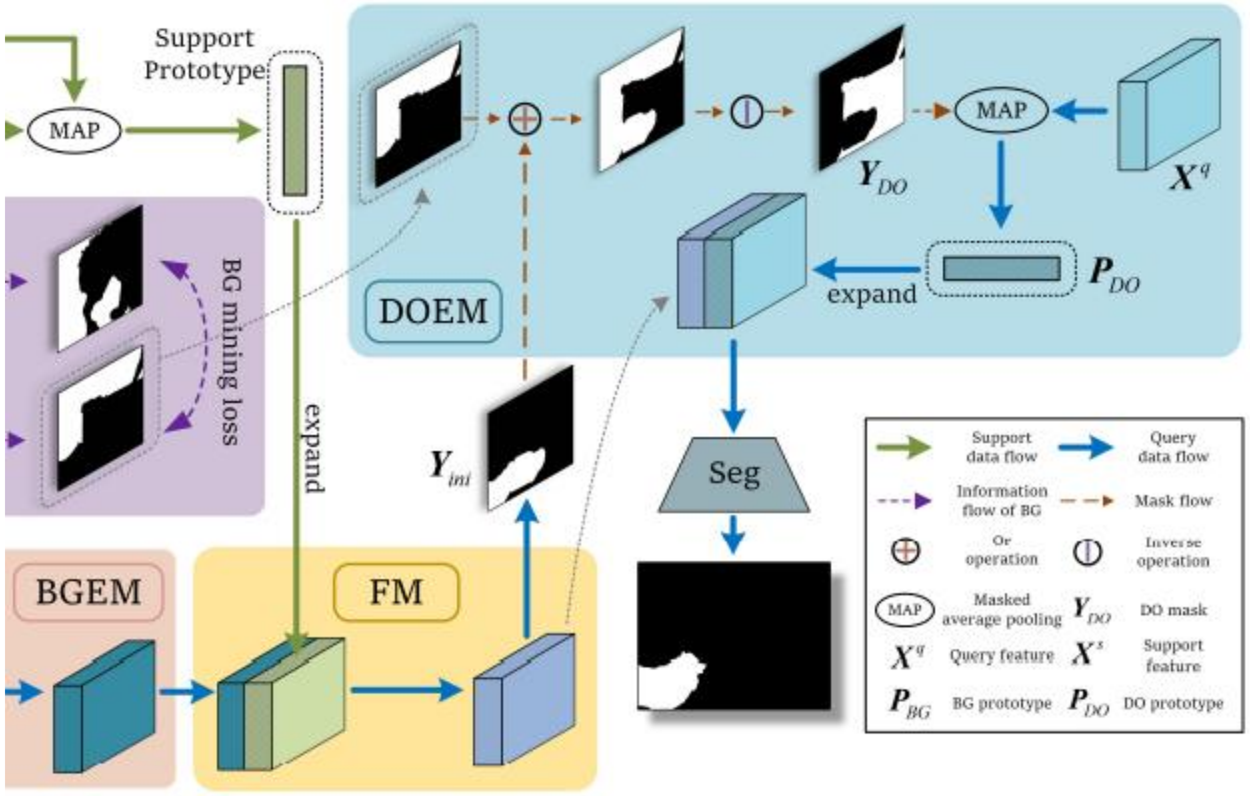
# Framework

DOEM:

将DO的预测表达为 BG与Target的补集

1. $Y_{DO}^{q} = 1 - (Y_{BG}^{q} \cup Y_{\mathrm{ini}}^{q})$
2. 用$Y_{DO}^{q}$作用在查询特征上 用MAP操作获得一个查询原型$P_{DO}$
3. 然后将$P_{DO}$展开为H×W×C 与激活的查询特征图 $X_{\mathrm{act}}^{q}$连接
4. 将其通过一个分割头得到最终结果

$$y^{q} = \mathrm{Seg}(X_{ini}^{q} \oplus \hat{P}_{DO}^{q}), \qquad (9)$$

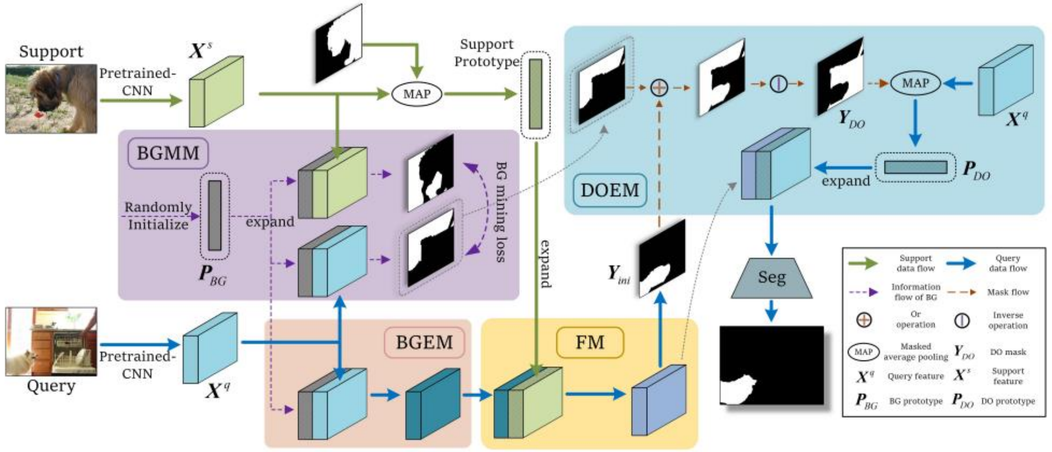


# Framework

Prototypical Contrastive Learning:

目的：希望使Taget与Dos之间的原型特征更具判别性 查询与支持的目标对象之间的原型更相似。

需要获得:
1.支持集的目标原型
2.DO原型（支持集和查询集）
3.查询集的目标原型

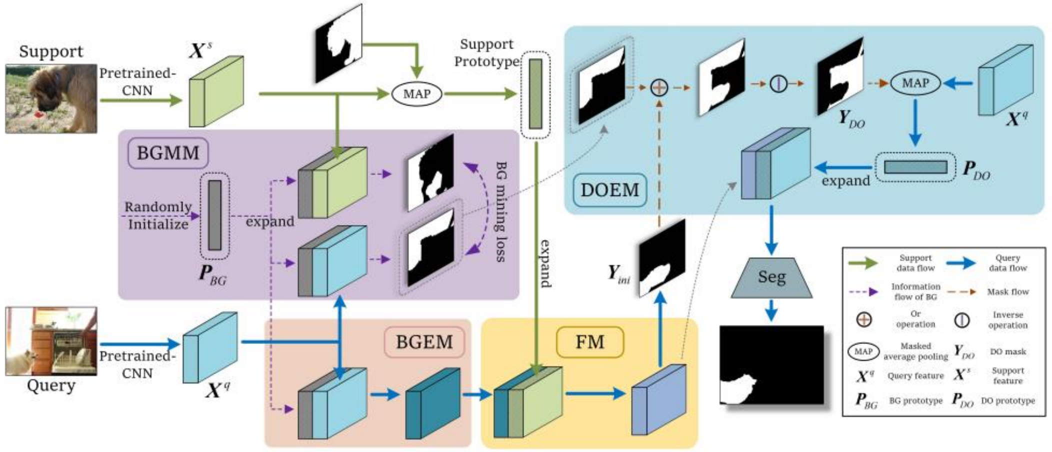


# Framework

Prototypical Contrastive Learning:

查询图像的目标原型$P_{\mathrm{q}}$
支持集的目标原型$P_{\mathrm{s}}$作为正样本
支持集和查询图像的DO原型均作为负样本

$$L_{PCL} = -log \frac{e^{cos(\boldsymbol{P}^q, \boldsymbol{P}^s)}}{\sum_{\mathcal{B}}\{e^{cos(\boldsymbol{P}^q, \boldsymbol{P}^q_{DO})} + e^{cos(\boldsymbol{P}^q, \boldsymbol{P}^s_{DO})}\}}, \quad (10)$$



# Results

Table 1. Class mIoU and FB-IoU results of four folds on PASCAL-$5^i$. The results of 'Mean' are the averaged class mIoU scores of all four folds. The detailed FB-IoU results of each fold are omitted in this table for simplicity. **Bold** indicates the best results.

| Backbone | Methods | 1-Shot | | | | | | 5-Shot | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Fold-0 | Fold-1 | Fold-2 | Fold-3 | **Mean** | FB-IoU | Fold-0 | Fold-1 | Fold-2 | Fold-3 | **Mean** | FB-IoU |
| VGG-16 | OSLSM [26] | 33.6 | 55.3 | 40.9 | 33.5 | 40.8 | 61.3 | 35.9 | 58.1 | 42.7 | 39.1 | 44.0 | 61.5 |
| | co-FCN [25] | 36.7 | 50.6 | 44.9 | 32.4 | 41.1 | 60.1 | 37.5 | 50.0 | 44.1 | 33.9 | 41.4 | 60.2 |
| | RPMM [37] | 47.1 | 65.8 | 50.6 | 48.5 | 53.0 | - | 50.0 | 66.5 | 51.9 | 47.6 | 54.0 | - |
| | PFENet [31] | 56.9 | **68.2** | 54.4 | 52.4 | 58.0 | 72.3 | 59.0 | **69.1** | 54.8 | 52.9 | 59.0 | 72.3 |
| | MMNet [35] | 57.1 | 67.2 | 56.6 | 52.3 | 58.3 | - | 56.6 | 66.7 | **63.6** | 56.5 | 58.3 | - |
| | NTRENet | **57.7** | 67.6 | **57.1** | **53.7** | **59.0** | **73.1** | **60.3** | 68.0 | 55.2 | **57.1** | **60.2** | **74.2** |
| ResNet-50 | CANet [42] | 52.5 | 65.9 | 51.3 | 51.9 | 55.4 | 66.2 | 55.5 | 67.8 | 51.9 | 53.2 | 57.1 | 69.6 |
| | RPMM [37] | 55.2 | 66.9 | 52.6 | 50.7 | 56.3 | - | 56.3 | 67.3 | 54.5 | 51.0 | 57.3 | - |
| | PFENet [31] | 61.7 | 69.5 | 55.4 | 56.3 | 60.8 | 73.3 | 63.1 | 70.7 | 55.8 | 57.9 | 61.9 | 73.9 |
| | SCL [41] | 63.0 | 70.0 | 56.5 | 57.7 | 61.8 | 71.9 | 64.5 | 70.9 | 57.3 | 58.7 | 62.9 | 72.8 |
| | ASGNet [14] | 58.8 | 67.9 | 56.8 | 53.7 | 59.3 | 69.2 | 63.7 | 70.6 | 64.2 | 57.4 | 63.9 | 74.2 |
| | ReRPI [1] | 59.8 | 68.3 | 62.1 | 48.5 | 59.7 | - | 64.6 | 71.4 | 71.1 | 59.3 | **66.6** | - |
| | SAGNN [36] | 64.7 | 69.6 | 57.0 | 57.2 | 62.1 | 73.2 | 64.9 | 70.0 | 57.0 | 59.3 | 62.8 | 73.3 |
| | MLC [38] | 59.2 | 71.2 | **65.6** | 52.5 | 62.1 | - | 63.5 | 71.6 | **71.2** | 58.1 | 66.1 | - |
| | NTRENet | **65.4** | **72.3** | 59.4 | **59.8** | **64.2** | **77.0** | **66.2** | **72.8** | 61.7 | **62.2** | 65.7 | **78.4** |
| ResNet-101 | DAN [32] | 54.7 | 68.6 | 57.8 | 51.6 | 58.2 | 71.9 | 57.9 | 69.0 | 60.1 | 54.9 | 60.5 | 72.3 |
| | PPNet [20] | 52.7 | 62.8 | 57.4 | 47.7 | 55.2 | 70.9 | 60.3 | 70.0 | 69.4 | 60.7 | 65.1 | 77.5 |
| | PFENet [31] | 60.5 | 69.4 | 54.4 | 55.9 | 60.1 | 72.9 | 62.8 | 70.4 | 54.9 | 57.6 | 61.4 | 73.5 |
| | ASGNet [14] | 59.8 | 67.4 | 55.6 | 54.4 | 59.3 | 71.7 | 64.6 | 71.3 | 64.2 | 57.3 | 64.4 | 75.2 |
| | ReRPI [1] | 59.6 | 68.6 | **62.2** | 47.2 | 59.4 | - | 66.2 | 71.4 | 67.0 | 57.7 | 65.6 | - |
| | MLC [38] | 60.8 | 71.3 | 61.5 | 56.9 | 62.6 | - | 65.8 | **74.9** | **71.4** | 63.1 | **68.8** | - |
| | NTRENet | **65.5** | **71.8** | 59.1 | **58.3** | **63.7** | **75.3** | **67.9** | 73.2 | 60.1 | **66.8** | 67.0 | **78.2** |

# Results

Table 3. Ablation study of the key modules in our NERTNet. mIoU results are reported on the PASCAL-$5^i$ dataset under the 1-shot setting.

| BGEM | DOEM | PCL | Fold-0 | Fold-1 | Fold-2 | Fold-3 | **Mean** | Precision |
|---|---|---|---|---|---|---|---|---|
| | | | 60.8 | 68.2 | 55.4 | 55.3 | 60.0 | 61.9 |
| ✓ | | | 63.2 | 71.1 | 57.7 | 57.4 | 62.4 | 62.8 |
| ✓ | ✓ | | 64.7 | 71.9 | 58.8 | 59.0 | 63.6 | 63.3 |
| ✓ | ✓ | ✓ | **65.4** | **72.3** | **59.4** | **59.8** | **64.2** | **63.6** |

# Results

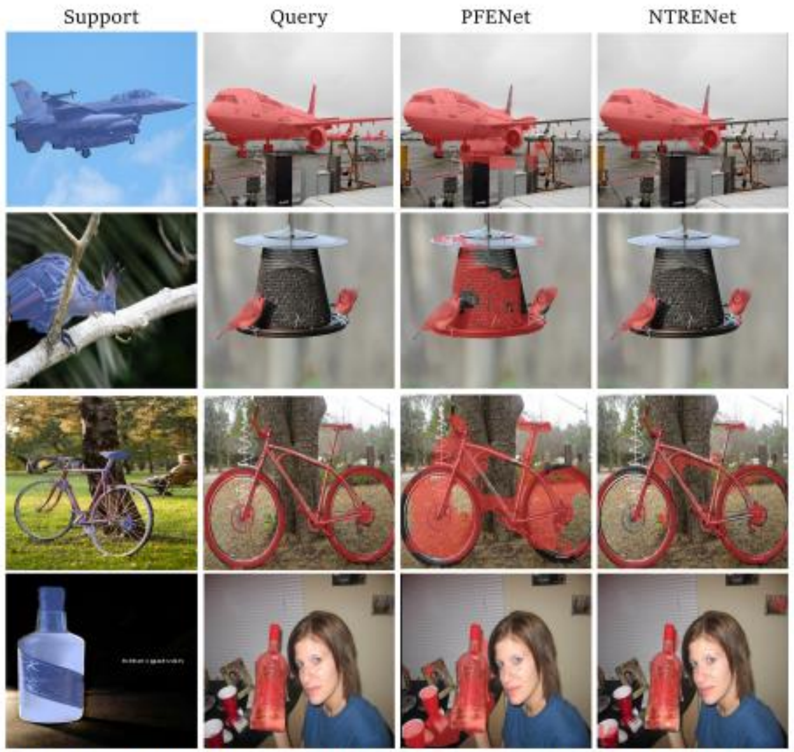



Figure 4. Qualitative results of our proposed NTRENet and PFENet. From left to right: support images, query images, prediction of PFENet, prediction of NTRENet.