| Method | Backbone | | | 1-shot | | | | | 5-shot | | | #learnable |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | split0 | split1 | split2 | split3 | mean | split0 | split1 | split2 | split3 | mean | params |
| PPNet ? | Res-50 | 52.7 | 62.8 | 57.4 | 47.7 | 55.2 | 60.3 | 70.0 | 69.4 | 60.7 | 65.1 | 31.5M |
| PMM ? | | 52.0 | 67.5 | 51.5 | 49.8 | 55.2 | 55.0 | 68.2 | 52.9 | 51.1 | 56.8 | - |
| PFENet ? | | 61.7 | 69.5 | 55.4 | 56.3 | 60.8 | 63.1 | 70.7 | 55.8 | 57.9 | 61.9 | 10.8M |
| CyCTR ? | | 65.7 | 71.0 | 59.5 | 59.7 | 64.0 | 69.3 | 73.5 | 63.8 | 63.5 | 67.5 | 7.4M |
| HSNet ? | | 64.3 | 70.7 | 60.3 | **60.5** | 64.0 | 70.3 | 73.2 | 67.4 | **67.1** | 69.5 | 2.6M |
| ASGNet ? | | 58.8 | 67.9 | 56.8 | 53.7 | 59.3 | 63.7 | 70.6 | 64.1 | 57.4 | 63.9 | 10.4M |
| SSP ? | | 61.4 | 67.2 | 65.4 | 49.7 | 60.9 | 68.0 | 72.0 | **74.8** | 60.2 | 68.8 | - |
| DCAMA ? | | 67.5 | 72.3 | 59.6 | 59.0 | 64.6 | **70.5** | 73.9 | 63.7 | 65.8 | 68.5 | - |
| **RiFeNet (Ours)** | | **68.4** | **73.5** | **67.1** | 59.4 | **67.1** | 70.0 | **74.7** | 69.4 | 64.2 | **69.6** | 7.7M |
| DAN ? | Res-101 | 54.7 | 68.6 | 57.8 | 51.6 | 58.2 | 57.9 | 69.0 | 60.1 | 54.9 | 60.5 | - |
| PMM ? | | 54.7 | 68.6 | 57.8 | 51.6 | 58.2 | 57.9 | 69.0 | 60.1 | 54.9 | 60.5 | - |
| PFENet ? | | 60.5 | 69.4 | 54.4 | 55.9 | 60.1 | 62.8 | 70.4 | 54.9 | 57.6 | 61.4 | 10.8M |
| CyCTR ? | | 67.2 | 71.1 | 57.6 | 59.0 | 63.7 | 71.0 | 75.0 | 58.5 | 65.0 | 67.4 | 7.4M |
| HSNet ? | | 67.3 | 72.3 | 62.0 | **63.1** | 66.2 | **71.8** | 74.4 | 67.0 | **68.3** | 70.4 | 2.6M |
| ASGNet ? | | 59.8 | 67.4 | 55.6 | 54.4 | 59.3 | 64.6 | 71.3 | 64.2 | 57.3 | 64.4 | 10.4M |
| SSP ? | | 63.7 | 70.1 | **66.7** | 55.4 | 64.0 | 70.3 | **76.3** | **77.8** | 65.5 | **72.5** | - |
| DCAMA ? | | 65.4 | 71.4 | 63.2 | 58.3 | 64.6 | 70.7 | 73.7 | 66.8 | 61.9 | 68.3 | - |
| **RiFeNet (Ours)** | | **68.9** | **73.8** | 66.2 | 60.3 | **67.3** | 70.4 | 74.5 | 68.3 | 63.4 | 69.2 | 7.7M |

Table 1: Performance comparison on PASCAL-$5^i$ in terms of mIoU (%).

| Method | Backbone | | | 1-shot | | | | | 5-shot | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | split0 | split1 | split2 | split3 | mean | split0 | split1 | split2 | split3 | mean |
| PPNet ? | Res-50 | 28.1 | 30.8 | 29.5 | 27.7 | 29.0 | 39.0 | 40.8 | 37.1 | 37.3 | 38.5 |
| PMM ? | | 29.3 | 34.8 | 27.1 | 27.3 | 29.6 | 33.0 | 40.6 | 30.3 | 33.3 | 34.3 |
| RPMMs ? | | 29.5 | 36.8 | 28.9 | 27.0 | 30.6 | 33.8 | 42.0 | 33.0 | 33.3 | 35.5 |
| CyCTR ? | | 38.9 | 43.0 | 39.6 | 39.8 | 40.3 | 41.1 | 48.9 | 45.2 | 47.0 | 45.6 |
| HSNet ? | | 36.3 | 43.1 | 38.7 | 38.7 | 39.2 | 43.3 | 51.3 | 48.2 | 45.0 | 46.9 |
| SSP ? | | **46.4** | 35.2 | 27.3 | 25.4 | 33.6 | **53.8** | 41.5 | 36.0 | 33.7 | 41.3 |
| DCAMA ? | | 41.9 | 45.1 | 44.4 | 41.7 | 43.3 | 45.9 | 50.5 | **50.7** | 46.0 | 48.3 |
| **RiFeNet (Ours)** | | 39.1 | **47.2** | **44.6** | **45.4** | **44.1** | 44.3 | **52.4** | 49.3 | **48.4** | **48.6** |

Table 2: Performance comparison on COCO in terms of mIoU (%).

the positive effects of the unlabeled branch decrease, with a proportional decline in performance gain in 5-shot.

Similar experiments on COCO support the above conclusion in Tab.2. Faced with a scenario with multiple objects in this dataset and a complex environment, RiFeNet still outperforms the current best DCAMA by 0.8% for almost all splits in the 1-shot setting. The comparison results demonstrate the benefits of RiFeNet. The unlabeled branch provides RiFeNet with richer relevant information, which in turn improves the performance of the model.

Qualitative results also prove the effectiveness of RiFeNet. In Fig.5 and Fig.A of the Appendix, the foreground objects in support and query images vary a lot, with inconsistent postures, appearances, and angles of photography. Despite this large intra-class variability, RiFeNet achieves significant improvement in maintaining foreground semantic consistency compared with the baseline. As for the similarity of background and foreground, the model deals with this binary identification much better even in cases with neighboring objects of similar appearance, with foreground occlusion, and with multiple classes of objects. Looking back to Fig.1, our extracted features are essential to maintain foreground semantic consistency and provide inter-class distinction for binary classification.

## 4.3 Ablation Studies

**RiFeNet improves the pixel-level binary classification.** To demonstrate the effectiveness of our proposed unlabeled enhancement and multi-level prototypes in RiFeNet, we conduct diagnostic experiments in Tab.3. All comparisons are set under a 1-shot setting, with ResNet50 as the backbone. Using either the unlabeled branch or multi-level prototype interaction results in a boost of approximately 2%. When two strategies work together, RiFeNet improves by 3.1% on top of the baseline.

**Different design choices of multi-level prototypes.** We conduct ablation experiments on the model design details mentioned for the multi-level prototype, as is shown in Tab.4. Consistent with the theoretical analysis in Sec.3.4, it proves that our practices such as adding guidance to unlabeled branches are reasonable and reliable.

**Different design choices of unlabeled branches.** We conduct experiments with different designed unlabeled branches to further explore their effect. As shown in Tab.5, the unlabeled branch without guided query prototypes results in even worse performance than the baseline, which

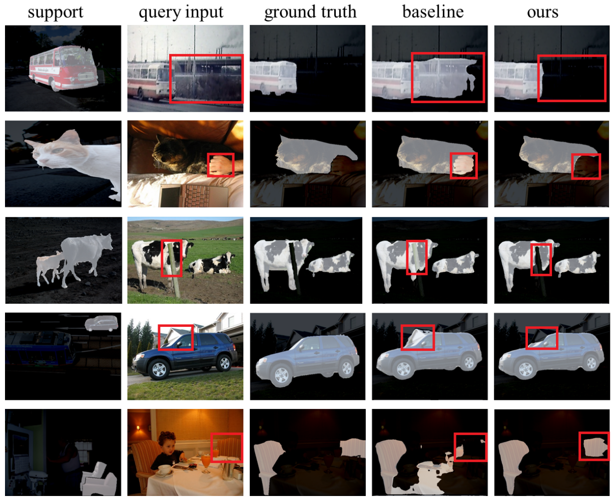

Figure 5: Qualitative segmentation results on novel classes on PASCAL-$5^i$. From *left* to *right*: support image with mask, query input, query ground-truth mask, query prediction of the baseline, and prediction of RiFeNet.

| Un | MP | split0 | split1 | split2 | split3 | mIoU |
|---|---|---|---|---|---|---|
| | | 65.7 | 71.0 | 59.5 | 59.7 | 64.0 |
| ✓ | | 67.3 | 71.8 | 66.2 | 59.2 | 66.1 |
| | ✓ | 66.0 | 72.1 | 66.2 | **60.4** | 66.2 |
| ✓ | ✓ | **68.4** | **73.5** | **67.1** | 59.4 | **67.1** |

Table 3: Ablation studies on the key components of RiFeNet. "Un" and "MP" denote the use of the unlabeled branch and the multi-level prototypes, respectively.

| components | split0 | split1 | split2 | split3 | mIoU |
|---|---|---|---|---|---|
| gp (support-only) | 67.3 | 71.8 | 66.2 | 59.2 | 66.1 |
| gp+gp | 67.5 | 73.1 | 66.2 | 58.4 | 66.3 |
| gp+lp (w/o CA) | 68.1 | 73.2 | 66.7 | 59.1 | 66.8 |
| gp+lp (w/ CA) | **68.4** | **73.5** | **67.1** | **59.4** | **67.1** |

Table 4: Ablation studies on multi-level prototypes. "gp" and "lp" denote global and local prototypes, respectively. That is, "gp+gp" means extracting both query and support prototypes globally. "CA" refers to channel-wise attention.

is consistent with our analysis in Sec.3.3. On the other hand, because the unlabeled inputs come from resampling the training dataset, we double the training iterations of the baseline for a fair comparison. Increased training iterations have little effect on the baseline due to early convergence. This proves that the effectiveness of our method is not from the multiple sampling of data but from the learned discriminative and semantic features.

**Different hyper-parameters.** We first look into the effect of different numbers of unlabeled input in a single meta-training process. Tab.6 shows the results on $PASCAL-5^i$ under a 1-shot setting, with ResNet50 as its backbone. The best results are obtained when the number of unlabeled images is set to 2. Initially, the segmentation effect of the model increased as the number of unlabeled images increased. When the number continues to increase, the accuracy decreases instead. We deem the reason is that when the effect of unlabeled enhancement counts much more than the query branch itself, the attention of feature mining may turn to the unlabeled branch, thus disturbing the query prediction. The segmentation accuracy decreases after the features are blurred. We also conduct detailed ablation experiments with other parameters, which are included in the Appendix.

| components | epoch | split0 | split1 | split2 | split3 | mIoU |
|---|---|---|---|---|---|---|
| w/o unlabel | 200 | 66.0 | 72.1 | 66.2 | **60.4** | 66.2 |
| w/o unlabel | 400 | 66.5 | 72.4 | 65.5 | 59.5 | 66.0 |
| un (w/o guide) | 200 | 66.9 | 72.2 | 65.9 | 58.3 | 65.8 |
| un (w/ guide) | 200 | **68.4** | **73.5** | **67.1** | 59.4 | **67.1** |

Table 5: Ablation studies on the unlabeled branch. "w/ guide" refers to the use of query local prototypes in the unlabeled branch for guidance, while "w/o guide" means using prototypes generated from the unlabeled branch itself.

| num | split0 | split1 | split2 | split3 | mIoU |
|---|---|---|---|---|---|
| 0 | 66.0 | 72.1 | 66.2 | 60.4 | 66.2 |
| 1 | 66.8 | 72.8 | 66.9 | 59.8 | 66.6 |
| 2 | **68.4** | **73.5** | **67.1** | 59.4 | **67.1** |
| 3 | 65.9 | 72.6 | 66.9 | **59.8** | 66.0 |

Table 6: Ablation studies of different numbers of unlabeled images in the single meta-training process.

## 5 Conclusion

In few-shot segmentation, traditional methods suffer from semantic ambiguity and inter-class similarity. Thus from the perspective of pixel-level binary classification, we propose RiFeNet, an effective model with an unlabeled branch constraining foreground semantic consistency. Without extra data, this unlabeled branch improves the in-class generalization of the foreground. Moreover, we propose to further enhance the discrimination of background and foreground by a multi-level prototype generation and interaction module. Rerum doloribus voluptatem, quibusdam aperiam mollitia culpa?Maiores ipsa quasi, hic aut laborum neque dicta maiores vel?Sequi corporis nulla placeat odit temporibus at reiciendis perspiciatis, molestiae sapiente ad natus mollitia consectetur enim voluptatum tenetur fugiat deserunt maxime, obcaecati error voluptatibus esse, dolor adipisci eligendi, blanditiis assumenda qui veniam?Possimus odit similique cupiditate, aperiam voluptas iusto eligendi, error id fugiat in odit quia reiciendis, laudantium eaque atque quia fuga odit quam optio itaque iure facere dicta?Ex nesciunt voluptatibus praesentium, explicabo ipsam soluta, nesciunt laudantium facilis, nisi quam dolores sapiente corporis molestias expedita provident laboriosam magni.Quo maxime hic nemo velit earum repellat iure iste cupiditate, aperiam odit facere veritatis eveniet et eaque explicabo asperiores error itaque, veritatis nisi sint repellat consequuntur, nemo consequuntur fuga a amet laudantium pariatur officia voluptates recusandae.Aspernatur ea accusantium vel eum modi, voluptas sapiente magni?