Read-only Prompt Optimization for Vision-Language Few-shot Learning (Supplementary Materials)

Here, we provide additional experiments and analyses which are not included in the main paper due to the limited space, which include ablation results of individual datasets and analysis about robustness of RPO on diverse benchmark datasets.

A. Ablation results on individual datasets

In Section 4.3 of our main paper, we only provided averaged results due to limited space. Therefore, here we provide analysis results for each of the 11 image recognition datasets.

A.1. Ablation on our components

Ablation on read-only mask As shown in Table 1, removing the read-only mask results in a drop of generalization performance on 10 out of 11 datasets based on the harmonic mean of base and novel accuracy. Additionally, novel accuracy deteriorates on all datasets without masked attention, which supports our claim that the read-only mechanism aids generalization.

Ablation on initialization Omitting the proposed initialization results in the improvement of 'base' accuracy and the loss of 'novel' accuracy on 7 out of 11 datasets. It shows worse generalization performance (evaluated by harmonic mean) on 9 out of 11 datasets. We believe that initializing prompts with pre-trained special tokens enables converging to a hypothesis with better generalizability.

A.2. Ablation on multi-modality of prompt

To show that the performance boost of RPO is not due to the multi-modality of prompts but the read-only mechanism, we provide the base-to-new generalization performance of RPO with only learnable text prompts (text-RPO). For text-RPO, we exploit global image feature of pre-trained ViT-B/16 backbone as image feature.

RPO with only learnable text prompts As shown in Table 2, text-RPO results in a marginal performance drop (0.8%) compared to RPO on average. To be specific, text-RPO even outperforms RPO on FGVCAircraft datasets by a small margin. Addtionally, text-RPO outperforms CoCoOp on 8 out of 11 datasets, which supports the effectiveness of read-only mechanism compared to conventional prompt tuning.

Table 1: **Ablation on our components in base to new generalization setting.** We evaluate base to new generalization performance by ablating read-only mask and special initialization.

(a) Average over 11 datasets

Methods	Base	Novel	H
w.o mask	79.97	69.90	74.02
w.o init	82.27	72.78	76.78
RPO	81.13	75.00	**77.78**

(b) ImageNet.

Methods	Base	Novel	H
w.o mask	76.10	69.03	72.40
w.o init	76.57	70.57	73.45
RPO	76.60	71.57	**74.00**

(c) Caltech101.

Methods	Base	Novel	H
w.o mask	98.10	92.87	95.41
w.o init	98.60	92.47	95.43
RPO	97.97	94.37	**96.03**

(d) OxfordPets.

Methods	Base	Novel	H
w.o mask	95.13	96.77	95.95
w.o init	95.57	96.90	**96.23**
RPO	94.63	97.50	96.05

(e) StanfordCars.

Methods	Base	Novel	H
w.o mask	74.53	71.47	72.96
w.o init	72.93	73.33	73.13
RPO	73.87	75.53	**74.69**

(f) Flowers102.

Methods	Base	Novel	H
w.o mask	96.87	71.47	82.92
w.o init	94.70	74.83	83.60
RPO	94.13	76.67	**84.50**

(g) Food101.

Methods	Base	Novel	H
w.o mask	88.97	88.73	88.85
w.o init	90.23	89.87	90.05
RPO	90.33	90.83	**90.58**

(h) FGVCAircraft.

Methods	Base	Novel	H
w.o mask	35.53	29.90	32.45
w.o init	38.37	31.43	34.55
RPO	37.33	34.20	**35.70**

(i) SUN397.

Methods	Base	Novel	H
w.o mask	78.96	76.10	77.50
w.o init	80.53	77.60	79.04
RPO	80.60	77.80	**79.18**

(j) DTD.

Methods	Base	Novel	H
w.o mask	78.67	55.77	65.25
w.o init	79.93	63.13	**70.53**
RPO	76.70	62.13	68.61

(k) EuroSAT.

Methods	Base	Novel	H
w.o mask	74.70	45.13	54.57
w.o init	93.80	55.13	69.37
RPO	86.63	68.97	**76.79**

(l) UCF101.

Methods	Base	Novel	H
w.o mask	82.07	70.63	75.92
w.o init	83.77	75.27	79.29
RPO	83.67	75.43	**79.34**

Table 2: **Comparison of CoCoOp, vis-RPO, text-RPO and RPO in the Base to new generalization setting.** We evaluate base to new generalization performance with unimodal RPO trained with 16-shot sampled training data.

(a) Average over 11 datasets

Methods	Base	Novel	H
CoCoOp	80.47	71.69	75.83
text-RPO	79.54	74.83	77.01
RPO	81.13	75.00	**77.78**

(b) ImageNet.

Methods	Base	Novel	H
CoCoOp	75.98	70.43	73.10
text-RPO	76.13	70.70	73.31
RPO	76.60	71.57	**74.00**

(c) Caltech101.

Methods	Base	Novel	H
CoCoOp	97.96	93.81	95.84
text-RPO	97.76	93.96	95.82
RPO	97.97	94.37	**96.03**

(d) OxfordPets.

Methods	Base	Novel	H
CoCoOp	95.20	97.69	**96.43**
text-RPO	94.56	97.16	95.84
RPO	94.63	97.50	96.05

(e) StanfordCars.

Methods	Base	Novel	H
CoCoOp	70.49	73.59	72.01
text-RPO	72.16	74.4	73.26
RPO	73.87	75.53	**74.69**

(f) Flowers102.

Methods	Base	Novel	H
CoCoOp	94.87	71.75	81.71
text-RPO	91.70	74.90	82.45
RPO	94.13	76.67	**84.50**

(g) Food101.

Methods	Base	Novel	H
CoCoOp	90.70	91.29	**90.99**
text-RPO	89.90	90.06	89.98
RPO	90.33	90.83	90.58

(h) FGVCAircraft.

Methods	Base	Novel	H
CoCoOp	33.41	23.71	27.74
text-RPO	36.5	35.7	**36.09**
RPO	37.33	34.20	35.70

(i) SUN397.

Methods	Base	Novel	H
CoCoOp	79.74	76.86	78.27
text-RPO	79.56	77.03	78.27
RPO	80.60	77.80	**79.18**

(j) DTD.

Methods	Base	Novel	H
CoCoOp	77.01	56.00	64.85
text-RPO	75.06	62.60	68.26
RPO	76.70	62.13	**68.61**

(k) EuroSAT.

Methods	Base	Novel	H
CoCoOp	87.49	60.04	71.21
text-RPO	79.90	71.60	75.52
RPO	86.63	68.97	**76.79**

(l) UCF101.

Methods	Base	Novel	H
CoCoOp	82.33	73.45	77.64
text-RPO	81.70	75.1	78.26
RPO	83.67	75.43	**79.34**