Towards Complete Activation: Foreground-Background Multi-Perspective Guided Cross-Support for Few-Shot Segmentation

Supplementary Material

1. Analysis of Multi-Prompt Design

To obtain more robust foreground-background localization, we extend the conventional single class-name prompt to a diverse set of foreground and background prompts, as illustrated in Figure 1(a). The motivation is that a single class-level description often captures only limited semantics and is insufficient to cover substantial intra-class variation. In contrast, multiple complementary prompts describe the target category from different semantic perspectives, thereby providing richer and more stable semantic guidance for prior generation.

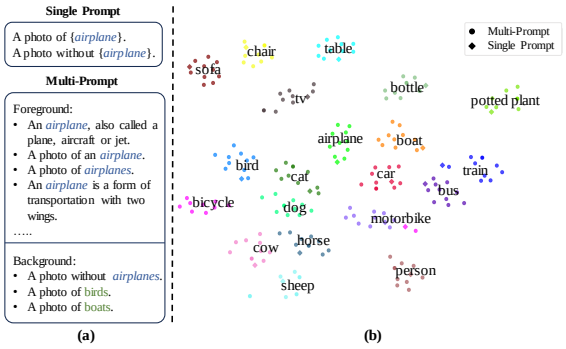

Figure 1. Comparisons of different prompt strategies. (a) Illustration of our multi-prompt design compared with the conventional single class-level prompt. (b) t-SNE visualization of prompt embeddings for 20 categories in PASCAL-5^i, showing that multi-prompt embeddings form more compact intra-class clusters and clearer inter-class separation.

To ensure prompt specification and reproducibility, our prompt design follows **class-agnostic FG/BG templates** together with **automated LLM-assisted generation** (Sec. 3.3.1 of the main paper). Specifically, we first construct template-based exemplars on five representative classes and refine them according to the quality of the resulting text priors. These fixed exemplars are then used to guide GPT-3.5 to automatically generate prompts for all remaining classes. Therefore, transferring the prompt design to a new dataset only requires the class names, while the same template set can be directly reused.

Figure 1(b) shows a t-SNE visualization of the prompt embeddings for the 20 categories in PASCAL-5^i. Compared with single-prompt representations, the proposed multi-prompt embeddings exhibit a more compact intra-class distribution and clearer inter-class separation, suggesting that they better capture the semantic diversity of each category while preserving category discriminability.

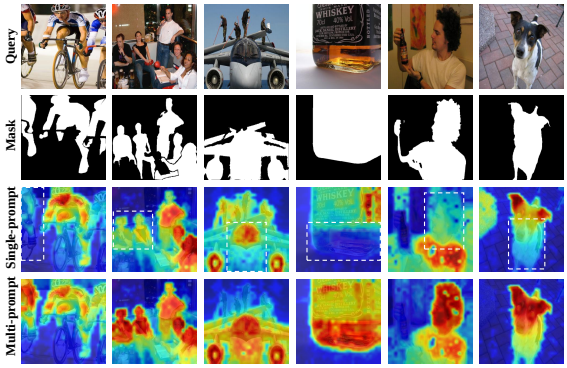

Figure 2. Qualitative comparison between the Single-Prompt and Multi-Prompt settings. From top to bottom are the query images, ground-truth (GT) masks, Single-Prompt priors, and Multi-Prompt priors, respectively. The white dashed box highlights regions that are easily overlooked.

Table 1. Ablation study on prompt count in terms of mIoU (%).

	Setting		Prompt Number					
No.	Varied	Fixed	1	3	5	7	10 (ours)	15
1	FG	BG=10	78.44	79.57	79.82	79.89	**80.06**	80.04
2	BG	FG=10	79.77	79.88	79.92	80.05	**80.06**	80.05

Table 2. Ablation study on prompt types. FG: class name (Cls.), synonyms (Syn.), grammatical variants (Gram.), definition (Def.), and appearance (App.). BG: exclusion (Excl.) and similar-category prompts (Sim.).

		FG					BG		Metrics	
No.	Setting	Cls.	Syn.	Gram.	Def.	App.	Excl.	Sim.	mIoU	FB-IoU
1	BaseSet	✓					✓		78.44	88.43
2	+Syn.	✓	✓				✓		79.47	89.25
3	+Gram.	✓		✓			✓		79.27	89.02
4	+Def.	✓			✓		✓		78.54	88.58
5	+App.	✓				✓	✓		79.09	88.92
6	FG-Full	✓	✓	✓	✓	✓	✓		79.77	89.50
7	BG-Sim.	✓	✓	✓	✓	✓		✓	79.37	89.29
8	Full	✓	✓	✓	✓	✓	✓	✓	**80.06**	**89.61**

Figure 2 further compares the activation priors generated by single-prompt and multi-prompt settings. Compared with the single-prompt setting, the proposed multi-prompt strategy produces more complete and spatially coherent activation maps. This advantage is particularly evident in challenging scenes containing multiple instances or large appearance variations, where richer semantic descriptions help the model better localize the target region.

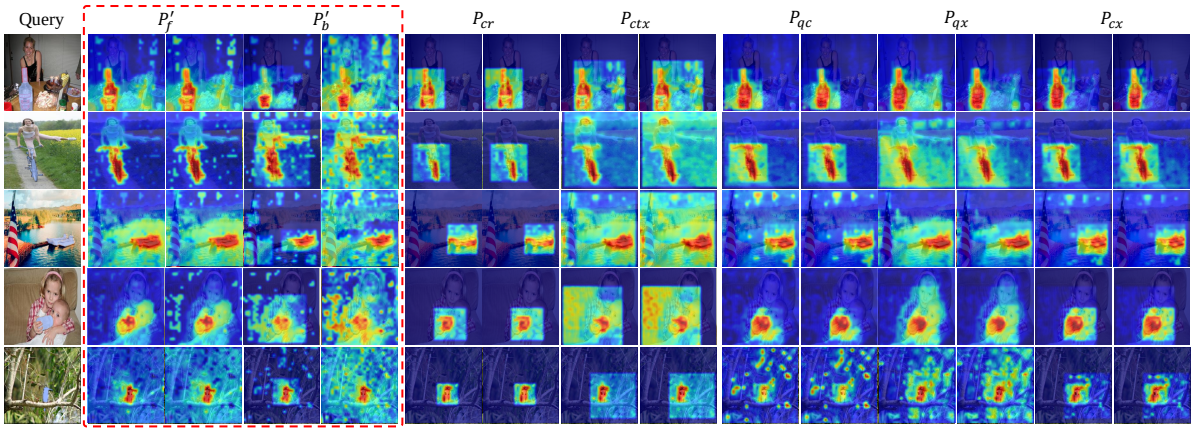

Figure 3. Qualitative results of the CRL refinement. Each type of prior contains two columns: the left column shows the CRL-refined prior, while the right column shows the original (unrefined) prior. The CRL operation yields more notable improvements for priors that suffer from strong background interference, such as P'_b.

Table 3. Ablation study of the Correlation-Refined Localization (CRL) operation on different priors. The symbol ✗ indicates that the corresponding prior is not refined by CRL.

	MFBL		I-Support		C-Support			mIoU	FB-IoU
No.	P'_f	P'_b	P_{cr}	P_{ctx}	P_{qc}	P_{qx}	P_{cx}	(%)	(%)
1								**80.06**	**89.61**
2	✗							79.94	89.53
3		✗						76.76	87.26
4	✗	✗						76.41	87.02
5			✗	✗				79.74	89.43
6					✗	✗	✗	79.64	89.30
7			✗	✗	✗	✗	✗	79.51	89.24

We further analyze the influence of prompt number in Table 1. As the number of prompts increases, the performance improves consistently for both the FG and BG branches, and gradually saturates around 10 prompts. This observation indicates that multiple prompts are beneficial because they enrich category semantics and improve prior robustness, while introducing more prompts beyond this point yields only marginal gains due to increasing redundancy. Based on this trade-off, we use 10 prompts in the final model.

Table 2 further studies the contribution of different prompt types. For the FG branch, synonym, grammatical-variant, and appearance prompts all improve over the basic class-name prompt, and combining all FG prompt types achieves better performance than using any single additional type alone. For the BG branch, exclusionary prompts and similar-category prompts provide complementary background semantics, and their joint use further improves performance. The best result is achieved when all FG and BG prompt types are jointly used, confirming that diverse prompt semantics are complementary and jointly beneficial for robust prior construction.

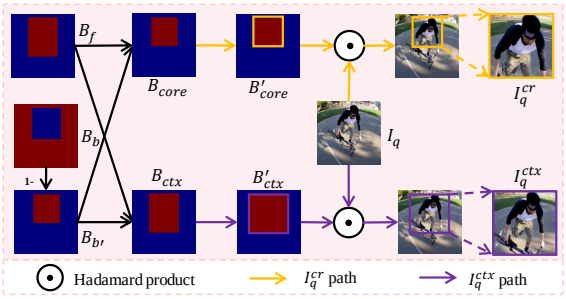

Figure 4. Structure of the AQT module. The core-focused region (B_{core}) and context-enriched region (B_{ctx}), derived from the localization boxes B_f and $B_{b'}$, are used to crop the query image I_q, producing the multi-perspective query variants (I_q^{cr}, I_q^{ctx}) where foreground regions are more prominently emphasized.

Table 4. Ablation study of different expansion strategies.

Expansion Method	mIoU(%)	FB-IoU(%)
Center	80.06	89.61
Top-Left	79.99	89.57
Bottom-Right	79.93	89.53

2. Analysis of Correlation-Refined Localization

We propose a Correlation-Refined Localization (CRL) operation that leverages internal correlations within query features to refine priors and reduce background interference. This operation is inserted at multiple stages of the model to enhance different types of priors. To thoroughly analyze its contribution, we evaluate the impact of CRL on each prior individually, as summarized in Table 3.

The results show that CRL yields the most significant improvement on the background prior P'_b in the MFBL

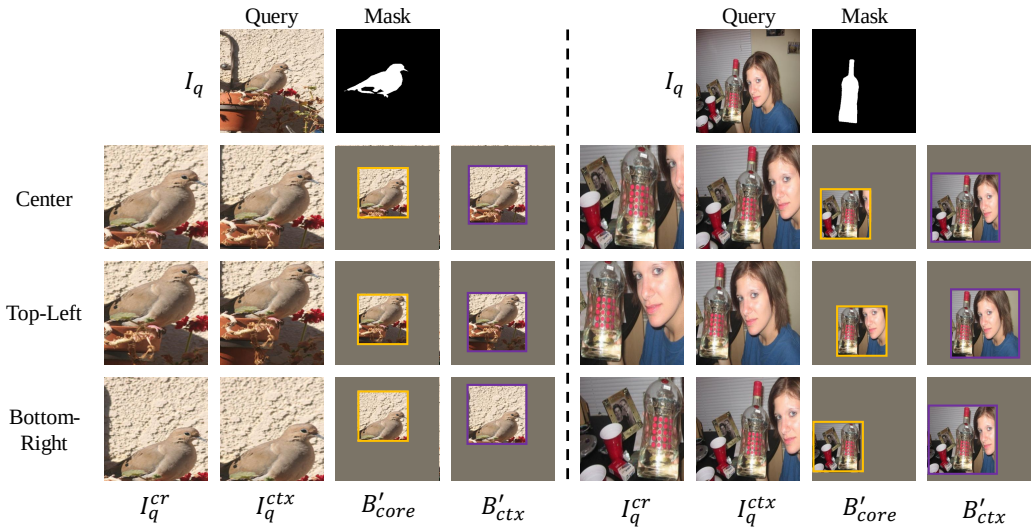

Figure 5. Qualitative comparison of different expansion and cropping strategies. The first row shows the original query images and their corresponding masks. The following three rows present the transformed images and their localization boxes obtained by expanding from the center, top-left, and bottom-right, respectively.

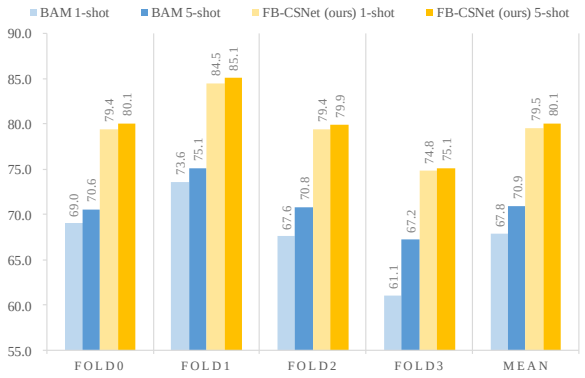

Figure 6. Comparison of the segmentation performance between BAM [1] (blue) and FB-CSNet (BAM) (yellow) on PASCAL-5^i.

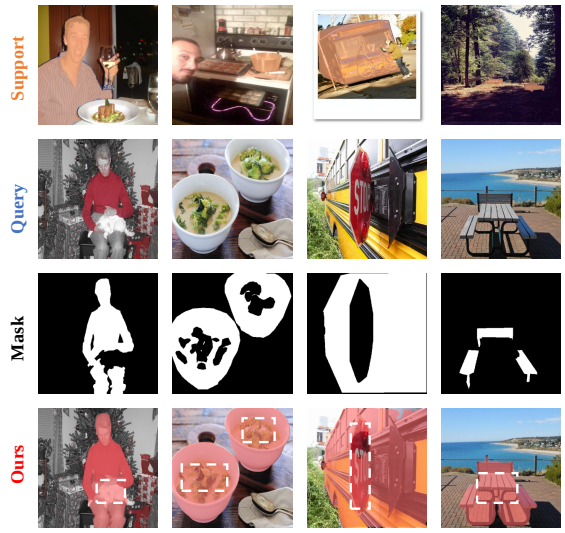

Figure 7. Representative failure cases. The white dashed boxes highlight over-segmented regions that may occur when the query image contains fine-grained or visually similar occlusions.

module. Background regions in query images are often cluttered, making background-derived priors prone to noise. By exploiting correlations within the semantically related regions, CRL suppresses such noise and produces cleaner background priors, which subsequently benefit localization and downstream cross-support interactions. In contrast, priors generated in the MPCS module (I-Support and C-Support) are already relatively stable, so CRL brings only marginal improvements in these branches.

The qualitative results in Figure 3 further support these findings. The refinement on the background prior P'_b is particularly prominent, as CRL effectively reduces spurious background responses and improves foreground activation. In summary, CRL is especially beneficial for priors that are more susceptible to background interference. To ensure optimal performance, we retain all effective CRL operations in the final version of our model.

3. Analysis of Adaptive Query Transformation

We introduce an Adaptive Query Transformation (AQT) mechanism to generate transformed query variants that emphasize target foreground regions. As shown in Figure 4, we derive two customized regions (B_{core}, B_{ctx}) from the fore-

ground and background localization boxes $(B_f, B_{b'})$ produced by the MFBL module, aiming to balance precise target localization and contextual coverage. Both regions are then expanded from their centers into square regions (B'_{core}, B'_{ctx}) and used to crop the original query image into the transformed variants I_q^{cr} and I_q^{ctx}. This step ensures consistent spatial alignment with the original query image while strengthening foreground emphasis.

Expanding regions of different shapes into squares ensures that the cropped images maintain the same spatial dimensions as the original input. We observe that different expansion strategies may lead to varying crop results and influence performance. Figure 5 compares three strategies: center-based expansion (first row), top-left expansion (second row), and bottom-right expansion (third row). As shown, center-based expansion best preserves the target object near the center and avoids truncating boundary details. The quantitative ablation in Table 4 further confirms that center-based expansion yields the best performance. Therefore, we adopt it as the default strategy in our final model.

4. Analysis of Generalization

To further examine the generalization capability of our approach, we replace the baseline HDMNet [2] with another representative FSS framework, BAM [1]. As shown in Figure 6, FB-CSNet (BAM) consistently and significantly outperforms the original BAM across all folds under both the 1-shot and 5-shot settings. This result demonstrates that our modules possess strong universality and robustness. They serve as plug-and-play components that effectively enhance different backbone networks without introducing additional trainable parameters.

5. Analysis of Failure Cases

Figure 7 presents representative failure cases. When the query image contains fine-grained or visually similar occlusions, the model may produce slightly over-segmented predictions. This behavior is likely related to the fixed confidence threshold used during high-confidence region selection for cross-support interaction, where small occluding regions may occasionally be included and subsequently propagated as foreground. These errors are highly localized and do not affect the overall effectiveness of our approach, but they do suggest room for further improvement. In future work, we plan to develop a more adaptive and discriminative confidence filtering strategy to better suppress misleading occlusion cues while preserving reliable foreground evidence.

References

[1] Chunbo Lang, Gong Cheng, Binfei Tu, and Junwei Han. Learning what not to segment: A new perspective on few-shot segmentation. In *CVPR*, pages 8057–8067, 2022. 3, 4

[2] Bohao Peng, Zhuotao Tian, Xiaoyang Wu, Chenyao Wang, Shu Liu, Jingyong Su, and Jiaya Jia. Hierarchical dense correlation distillation for few-shot segmentation. In *CVPR*, pages 23641–23651, 2023. 4