FCC: Fully Connected Correlation for One-Shot Segmentation

Supplementary Material

1. Additional Analysis

Our proposed method, FCC, demonstrates impressive performance on the PASCAL [6] and COCO [20] datasets. This success is achieved by leveraging both cross-layer and same-layer correlations to construct a comprehensive correlation map. This approach provides robust prior information, enabling the network to identify target features. In these supplementary materials, we present the following additional analyses.

1. Motivation for FCC
2. Compare with HSNet [26] and MSI [28]
3. Further analysis of the correlation between cross-level.
4. Scalability and deployment feasibility
5. Training efficiency
6. 5-shot validation with FB-IoU metric
7. Additional qualitative results on PASCAL and COCO.
8. Details of data splits for PASCAL, COCO, and Generalization test.
9. Analysis of failure cases.
10. Visualization of the entire heatmap of 1x1 conv weights.

1.1. Motivation for FCC

Figure 4 from the main paper presents CKA similarity heatmaps between support and query features for the same object under variations in scale, shape, and color. These heatmaps reveal that strong feature similarity does not arise solely from same-layer comparisons. Notably, when the object appears at a different scale or is partially occluded, the regions of high similarity shift away from the diagonal entries, where diagonal values represent same-layer similarity. Our ablation study in the main Tables 5 and 8 quantitatively demonstrates the benefit of incorporating cross-layer correlations, showing consistent performance gains over same-layer-only baselines. Table 7 from the main paper further explores the effect of using different combinations of layers, reinforcing that fully utilizing cross-layer correlations leads to the best performance.

1.2. Compare with HSNet and MSI

Both HSNet [26] and MSI [28] extract features from multiple layers and then compute correlation maps based solely on same-layer feature comparisons. These methods limit their correlation to features extracted at the same network depth (Main Fig. 2). In contrast, FCC is the first approach to compute both same-layer and cross-layer correlation maps in few-shot segmentation, which is intuitively motivated (Main Fig. 4) and extensively validated (Main Tables 1,5,8). This enables FCC to fully exploit the rich and complementary information distributed throughout the feature hierarchy

1.3. Correlation Analsis

We visualized the CKA similarity heatmap to identify patterns in correlation maps during cross-layer comparisons (Fig. 2). Interestingly, many correlation maps show high similarity scores (>0.7), even between features from different layers. A notable example is the first layer of query features and the last layer of support features, which exhibit a similarity score of 0.74. Additionally, all correlation maps were visualized using heatmaps, demonstrating that FCC effectively utilizes the diverse variations in correlation map information to identify target features.

Fig. 3 illustrates differences between ResNet50 and DINOv2 in feature correlation across layers. ResNet50 [8] exhibits a high concentration of features with minimal variation, whereas DINOv2 [30] shows more diverse correlations. This characteristic allows FCC to capture richer details of a target object, making it particularly effective for segmentation tasks involving varying object scales. Due to mismatched spatial and channel sizes, we could not compare all layers in ResNet, and black areas represent regions where correlations could not be calculated.

1.4. Scalability and deployment feasibility

FCC introduces an additional 77.3 GMac over the baseline while still achieving approximately 40 FPS. The increased computation is confined to the correlation calculation, and the parameter numbers grow by only 0.1M, adding just 3 MB to the baseline model size (Main Table 6).

1.5. Training efficiency

We found that FCC significantly enhanced training efficiency, achieving the best-performing model that is approximately 2.7 times faster than the baseline on both the PASCAL and COCO datasets (Fig. 1). This improvement is caused by the strong prior information that FCC provides, which enables the network to converge significantly faster. The detailed number of epochs can be found in Supplementary Material Table 1.

During training, only the FCC module is fine-tuned, with pretrained image backbones frozen to isolate their effect. This ensures efficiency gains stem solely from FCC's strong priors. FCC converges significantly faster than the baseline, using identical backbones for fair comparison. Additionally, this cross-layer correlation allows the model to find the most telling similarity. For instance, it can match a fine-grained texture detail (a low-level feature) in the support

Benchmark	5-shot					
	Fold^0	Fold^1	Fold^2	Fold^3	mIoU	FB-IoU
PASCAL	81.6	87.8	81.8	85.4	84.2	90.7
COCO	67.3	72.5	70.0	69.6	69.9	83.6

Table 1. 5-shot evaluation on PASCAL-5^i and COCO-20^i. FCC with DINOv2 used.

image to the overall shape of the object (a high-level feature) in the query image. This ability to make flexible, non-obvious connections makes the model extremely robust to changes in object scale, angle, or appearance, effectively making its prior knowledge more powerful and useful for the specific task.

1.6. 5-shot validation with FB-IoU metric

We performed a 5-shot evaluation using k-shot settings, where all predictions were aggregated and normalized based on the highest score. We report both mean Intersection over Union (mIoU) and Foreground-Background Intersection over Union (FB-IoU) as evaluation metrics. FB-IoU is a simplified variant of mIoU that treats all foreground classes as a single class, making it more robust to class imbalance and providing a clearer measure of general segmentation quality. (Table 1)

1.7. Additional Qualitative Results

FCC achieves higher mIoU scores than other methods across standard benchmarks, PASCAL [6] and COCO [20]. The qualitative examples in Fig.4 and Fig.5 illustrate the effectiveness of FCC, showcasing FCC's ability to accurately identify and segment target objects in query images.

1.8. Data Split

Table 2 shows the classes used for testing in each fold for PASCAL-5^i [6], COCO-20^i [20], and the Generalization test, following the methodology of previous works [26–28, 39]. Classes excluded from testing are used for training.

1.9. Failure Cases

Although FCC shows significant advancements over current state-of-the-art methods on PASCAL [6] and COCO [20] datasets, further improvements can be made by addressing the failure cases depicted in Fig. 6. The key challenges identified are as follows:

- Extreme differences between support and query objects: FCC encounters difficulties when the objects in the support set differ significantly from the target object in the query image.
- Severely limited target information in support set: FCC struggles to identify the target object in the query image when the available target information is drastically limited.
- Difficulty with small and thin object: The trained network struggles to segment very small and thin objects with clarity.

1.10. 1x1 Conv Weights

Fig. 7 shows how 1x1 convolution weights in FCC prioritize diverse correlations, effectively segmenting the target object by leveraging both same-level and cross-level correlations for comprehensive target information.

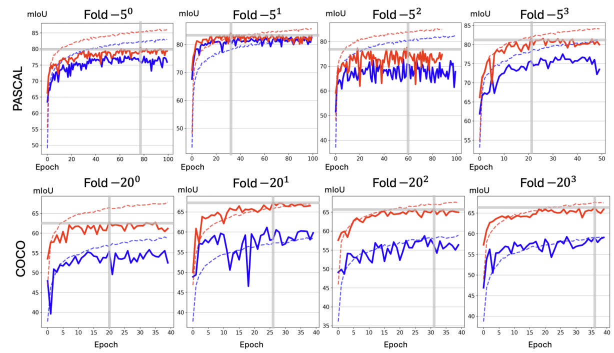

Figure 1. Training profiles of FCC on PASCAL-5^i [6] and COCO-20^i [20]. The gray line marks the point where FCC achieved the highest mIoU accuracy.

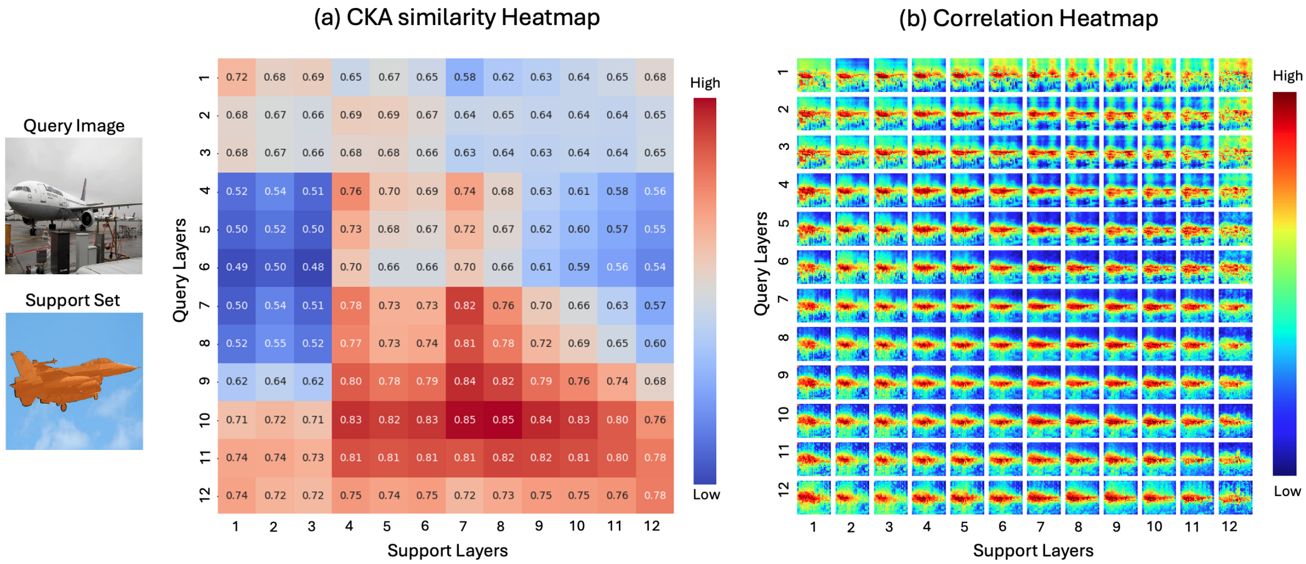

Figure 2. (a) The Centered Kernel Alignment (CKA) similarity heatmap between target and query features across different layers reveals that features from various layers capture different levels of local and global information about the target object. (b) Visualization of correlation map from different layers. Although the differences are not huge, each correlation map maintains subtle variations, showing that they are not identical and contain different information.

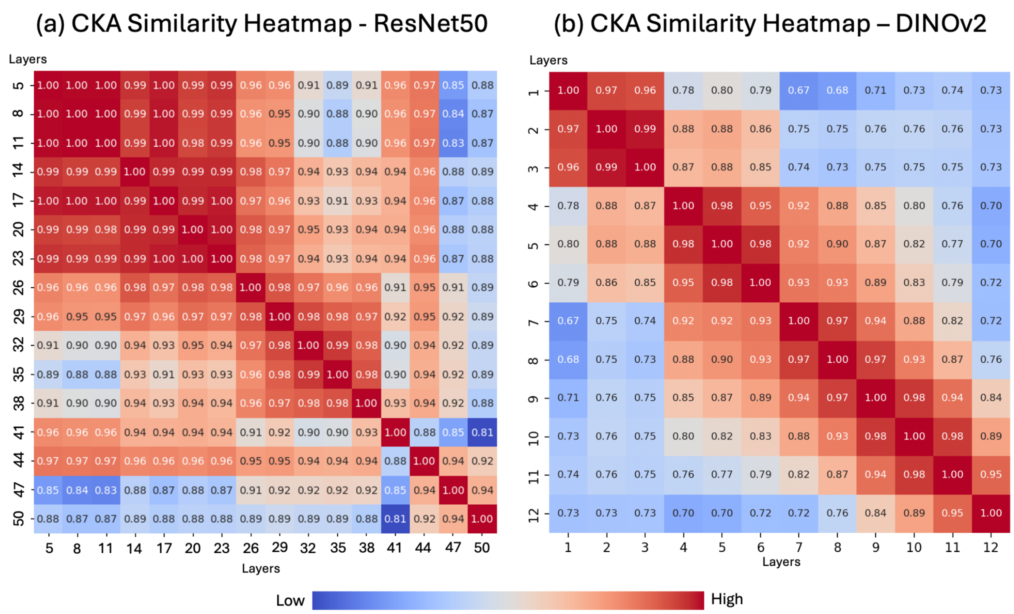

Figure 3. CKA Similarity Heatmap comparison between ResNet50 [8] and DINOv2 [30]. Features extracted from the same object. (a) shows high similarity among the early layers of the network. As the layers go deeper, the similarity gradually decreases. (b) displays a more uniform similarity distribution across layers. This suggests that DINOv2 [30] maintains a consistent representation across its layers.

Figure 4. FCC Qualitative result on PASCAL-5^i [6]

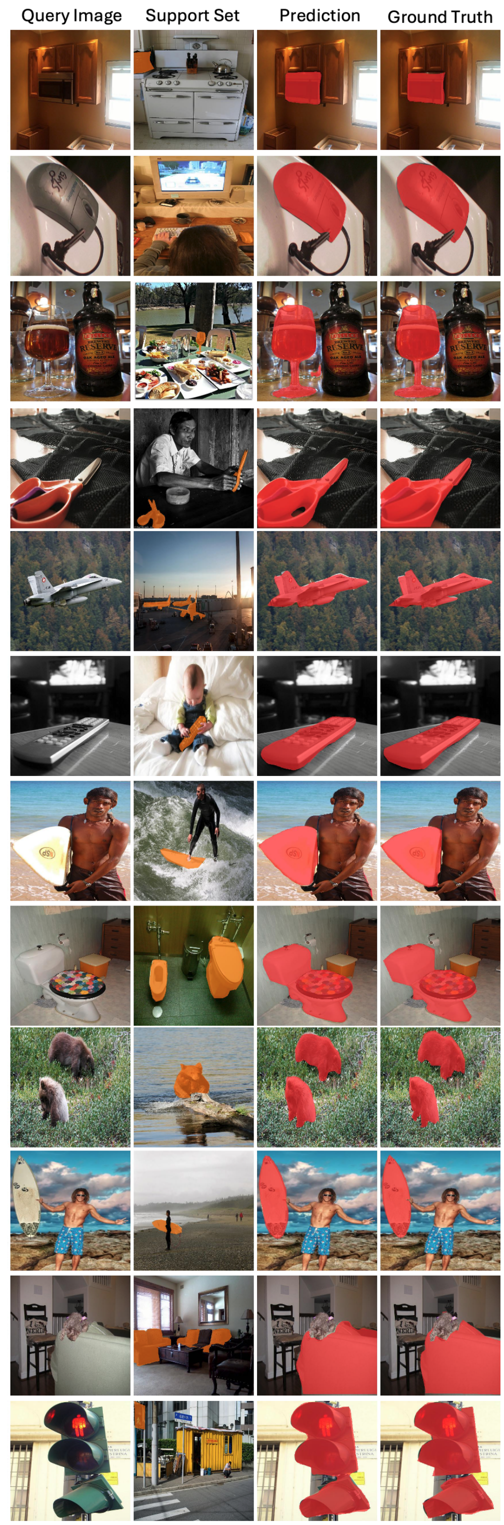

Figure 5. FCC Qualitative result on COCO-20^i [20]

Figure 6. FCC Failure cases on PASCAL-5^i [6] and COCO-20^i [20].

Fold	PASCAL-5^i Test Classes	COCO-20^i Test Classes	PASCAL-5^i Generalization Test Classes
0	Aeroplane, Bicycle, Bird, Boat, Bottle	Person, Airplane, Boat, Parking meter, Dog, Elephant, Backpack, Suitcase, Sports ball, Skateboard, Wine glass, Spoon, Sandwich, Hot dog, Chair, Dining table, Mouse, Microwave, Sink, Scissors	Aeroplane, Boat, Chair, Dining table, Dog, Person
1	Bus, Car, Cat, Chair, Cow	Bicycle, Bus, Traffic light, Bench, Horse, Bear, Umbrella, Frisbee, Kite, Surfboard, Cup, Bowl, Spoon, Orange, Pizza, Couch, Toilet, Remote, Oven, Book, Teddy bear	Horse, Sofa, Bicycle, Bus
2	Dining table, Dog, Horse, Motorbike, Person	Car, Train, Fire hydrant, Bird, Sheep, Zebra, Handbag, Skis, Baseball bat, Tennis racket, Fork, Banana, Broccoli, Donut, Potted plant, TV, Keyboard, Toaster, Clock, Hair drier	Bird, Car, Potted plant, Sheep, Train, TV/monitor
3	Potted plant, Sheep, Sofa, Train, TV/monitor	Motorcycle, Truck, Stop sign, Cat, Cow, Giraffe, Tie, Snowboard, Baseball glove, Bottle, Knife, Apple, Carrot, Cake, Bed, Laptop, Cell phone, Sink, Vase, Toothbrush	Bottle, Cow, Cat, Motorbike

Table 2. Details of the data splits for PASCAL-5^i, COCO-20^i, and Generalization test.

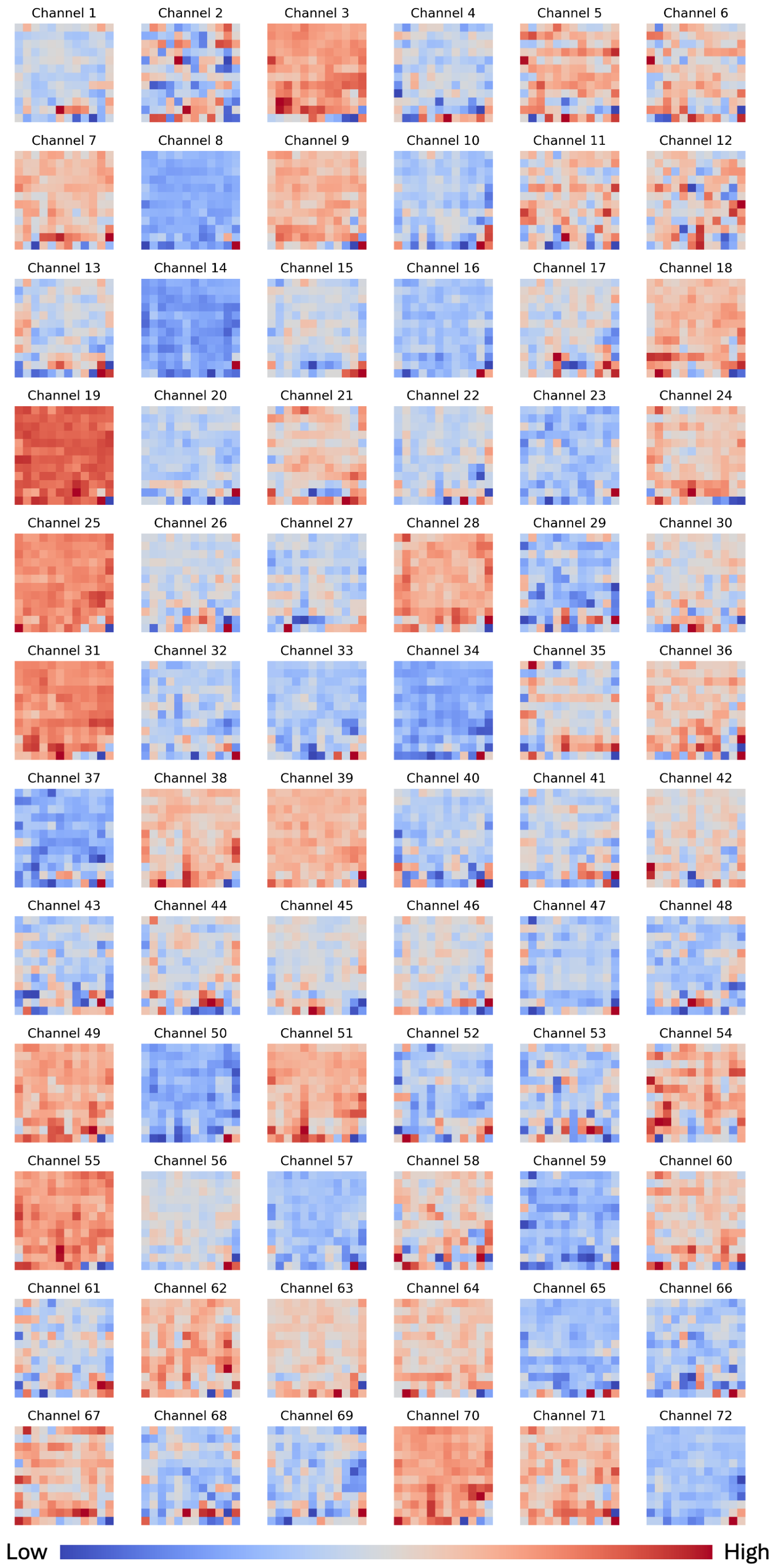

Figure 7. Heatmap of 1x1 convolution weights in FCC.