

# Supplementary Materials

Anonymous Authors

## 1 ABLATION STUDY ON THE LOCATION OF THE GROUPED STYLE MODULATION LAYER

As shown in Figure 1, we explore various embedding methods of the Group Style Modulation (GSM) layer at different positions within the network. We integrate the GSM layer before the residual connection (BI), after the residual connection (AI), and at the end of each network layer (AL). Additionally, we experiment with placing the GSM layer parallel to the CBR and CB layers (PI), directly incorporating the diverse features into the residual connections. Furthermore, we replace all BN layers with GSM layers (IB).

The specific experimental results are presented in Figure 2. From the data, BI and AL achieves similar performances, which can be attributed to both receiving inputs from the output of ReLU. ReLU transforms all negative values to zero, consequently narrowing the potential style range that GSM can learn, leading to reduced generalization capabilities. The results from IB suggest that retaining BN features contributes to improved performance. Compared to PI, our use of AI yield better prediction results, benefiting from the features extracted by the convolutional neural network.

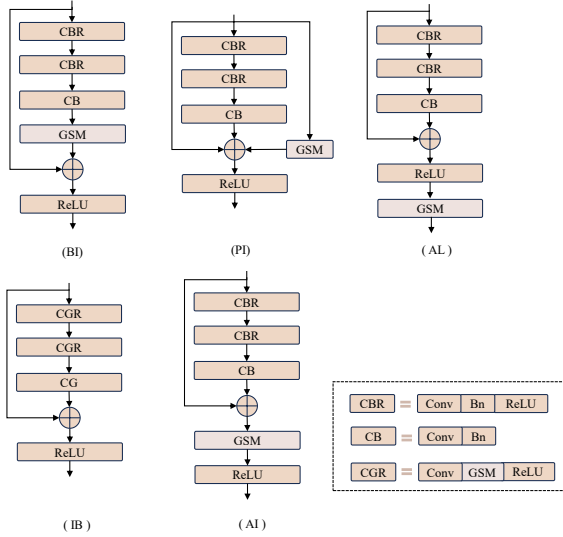


Figure 1: Various embedding methods of the group style modulation layer.

## 2 ADDITIONAL QUALITATIVE RESULTS

As illustrated in Figure 3, we demonstrate the visualization results using a dual-branch fusion approach. By weighted blending of the branches with and without the Group Style Modulation, our method achieves more accurate segmentation results. Specifically, we can observe that class-irrelevant noise typically appears in the prediction results of the branch with GSM, while the branch without GSM

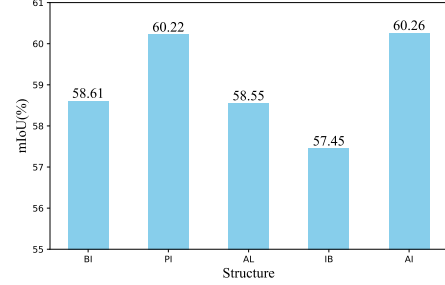


Figure 2: Prediction results for the location of the grouped style modulation layer.

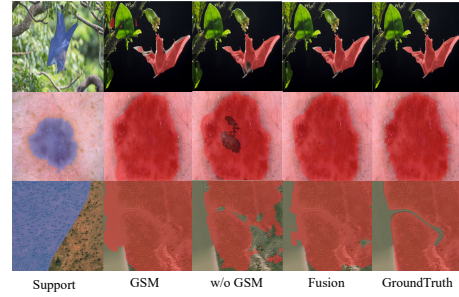


Figure 3: Illustration of the visual prediction results on query samples using branches with GSM, branches without GSM, and the method that uses dual-branch fusion.

tends to miss essential category information. For instance, in the second row, the non-GSM branch fails to predict the middle section, whereas the GSM branch predicts a broader area in the bottom left corner of the target. This occurs because, although GSM can effectively aid the model in learning diverse domain style features and enhance its generalization ability, it might also induce potential semantic drift, making it challenging to ensure the discriminative quality of segmentation results. Conversely, the branch without GSM focuses more on the category information itself. Our proposed dual-branch fusion strategy, which weights the predictions from the GSM and non-GSM branches, effectively enhances the robustness and discriminative power of the predictions.

Additionally, Figure 4 further presents the segmentation results of our method in the 5-shot setting, complementing the 1-shot segmentation results discussed in the main text.

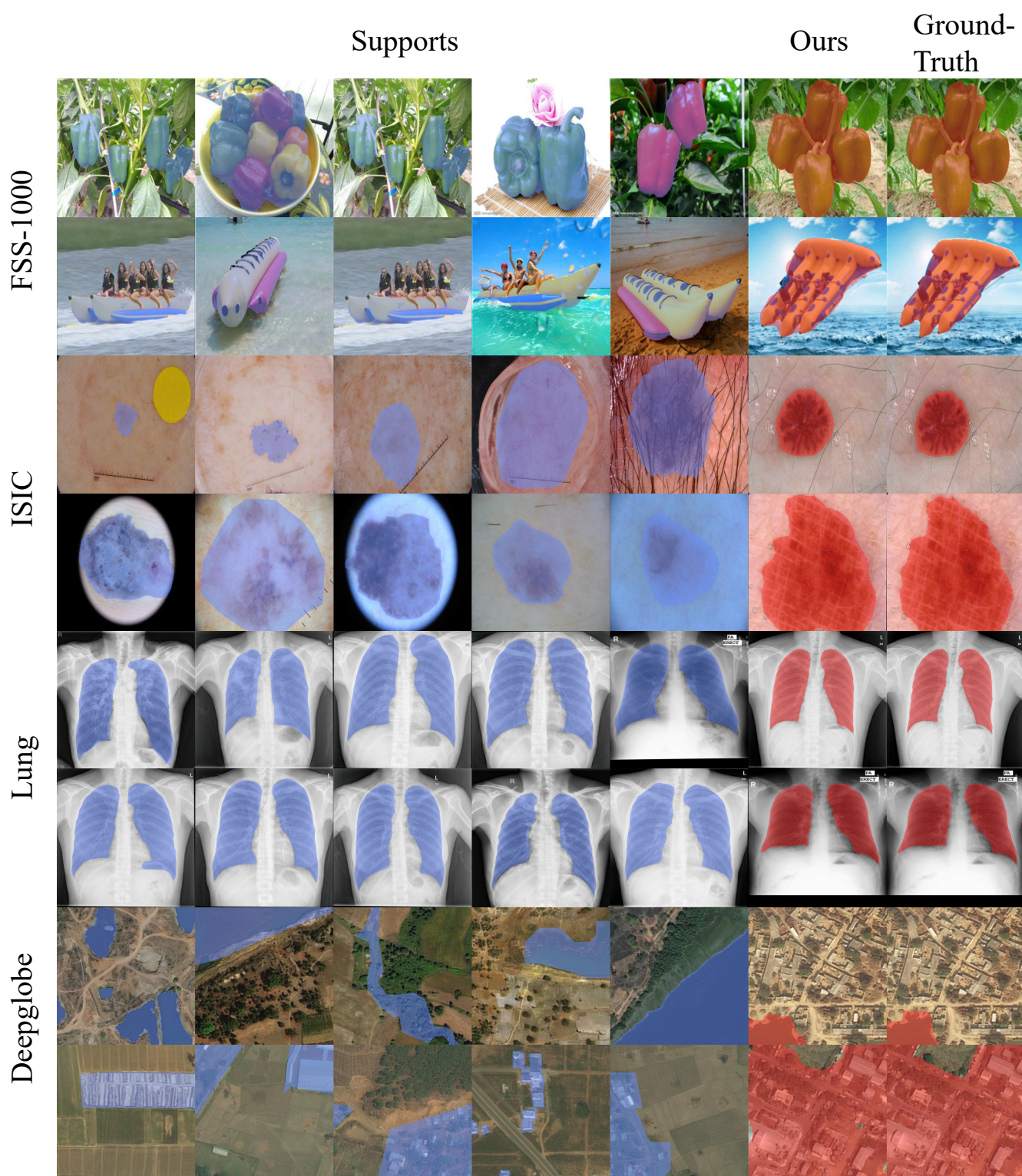


Figure 4: Prediction results of our model on the 5-shot setting for CD-FSS task. From top to bottom, each row represents FSS-1000, ISIC, Chest X-Ray, and Deepglobe datasets.