

MEGANet: Multi-Scale Edge-Guided Attention Network for Weak Boundary Polyp Segmentation (Supplementary Material)

Nhat-Tan Bui¹, Dinh-Hieu Hoang^{2,3}, Quang-Thuc Nguyen^{2,3}, Minh-Triet Tran^{2,3}, Ngan Le¹

¹University of Arkansas, Fayetteville, Arkansas, USA

²University of Science, and John von Neumann Institute, VNU-HCM

³Vietnam National University, Ho Chi Minh City, Vietnam

1. Visualization of the level-1 Laplacian pyramid

As previously mentioned, our Edge-Guided Attention (EGA) module takes three inputs, with one of them being the high-frequency feature denoted as f^l . In most cases, f^l corresponds to $L_1(I)$, where L_1 represents the first-level image of the Laplacian pyramid, and I is the input image. The visual representation of f^l can be seen in Figure 1.

Obviously, the boundaries between different elements in the image are highlighted. The nuances of the texture are also exposed, highlighting its ability to preserve the high-frequency information inherent in the Laplacian pyramid. Leveraging this capability, the high-frequency feature f^l effectively guides the model’s attention towards edges relevant to polyps, thus enhancing predictions at each decoder level.

Upon examining polyp images from the Kvasir dataset, we’ve come to realize that the mucous membrane (background) of this dataset is considerably intricate. Consider the final image in Figure 1 and Figure 2 as examples; distinguishing between the appearance of polyps and the surrounding colonic mucosa proves to be quite challenging, even for human observers. This complexity is why the high-frequency feature f^l extracted from the Kvasir dataset might contain noise, as it encompasses a significant amount of edge information from the mucous membrane. Consequently, this noise contributes to the scenario where our method, without the incorporation of f^l , achieves the highest score within the Kvasir dataset.

2. Visualization of the decoded predicted feature

Another input of our Edge-Guided Attention (EGA) module at certain layer i^{th} is the predicted feature map at a higher layer $(i + 1)^{th}$, denoted as \hat{f}_{i+1}^d . The \hat{f}_{i+1}^d is then decomposed into the reverse attention map \hat{f}_{i+1}^r and the

boundary attention map \hat{f}_{i+1}^b . Figure 3 depicts the heatmap of an example for \hat{f}_i^d , \hat{f}_i^b and \hat{f}_i^r during the decoding phase where $i = [1, 2, 3, 4]$. As we can see, the boundary attention \hat{f}_i^b , which is attained from the Laplacian operator, efficiently concentrates on the boundary between the polyp object and the surrounding mucous membrane. This feature map can force the lower decoder layer to focus on the boundary of the prediction map from the higher layer better. The reverse attention \hat{f}_i^r , also shown in Figure 3, guides the model to separate the polyp and mucous membranes by removing and refining the imprecise prediction map from the higher layer. By combining \hat{f}_i^b and \hat{f}_i^r , the model can have the ability to recognize the whole object regions and better distinguish polyp from the background by accurate boundary.

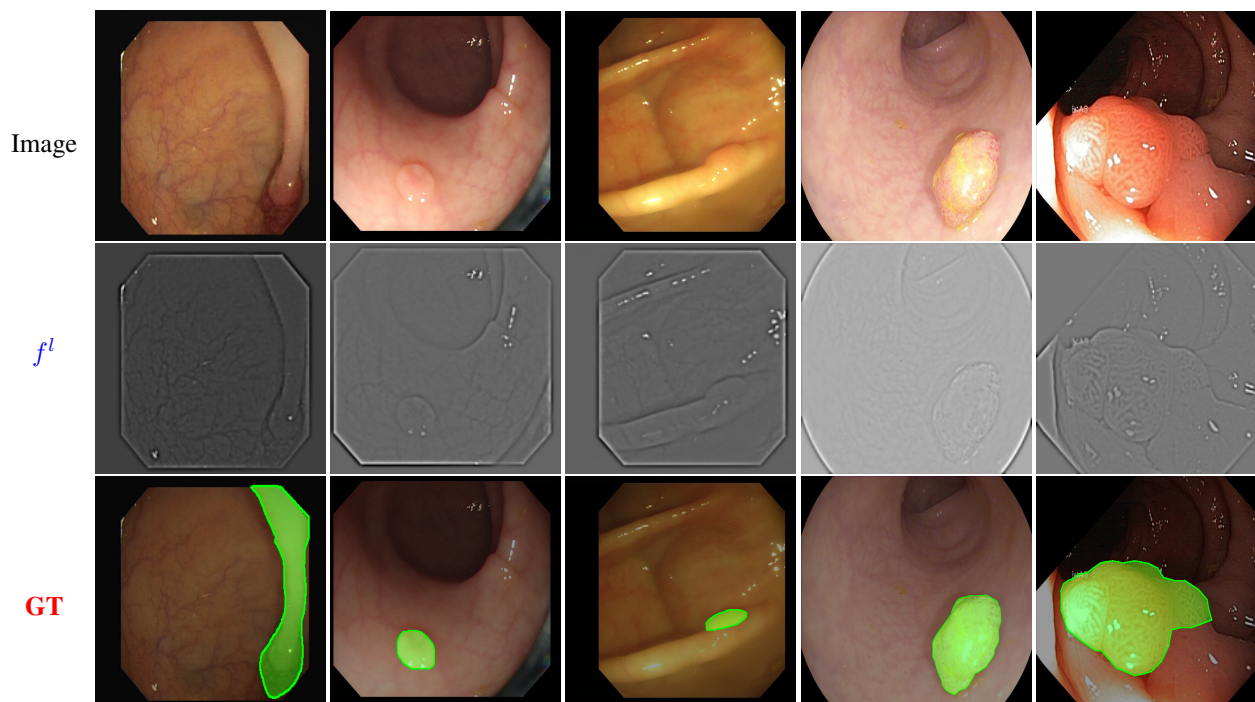


Figure 1. Visualization of the level-1 Laplacian pyramid in polyp images, which is basically the same with the high-frequency feature f^l . From left to right: the images come from the ClinicDB, ColonDB, CVC-300, ETIS and Kvasir datasets, respectively.

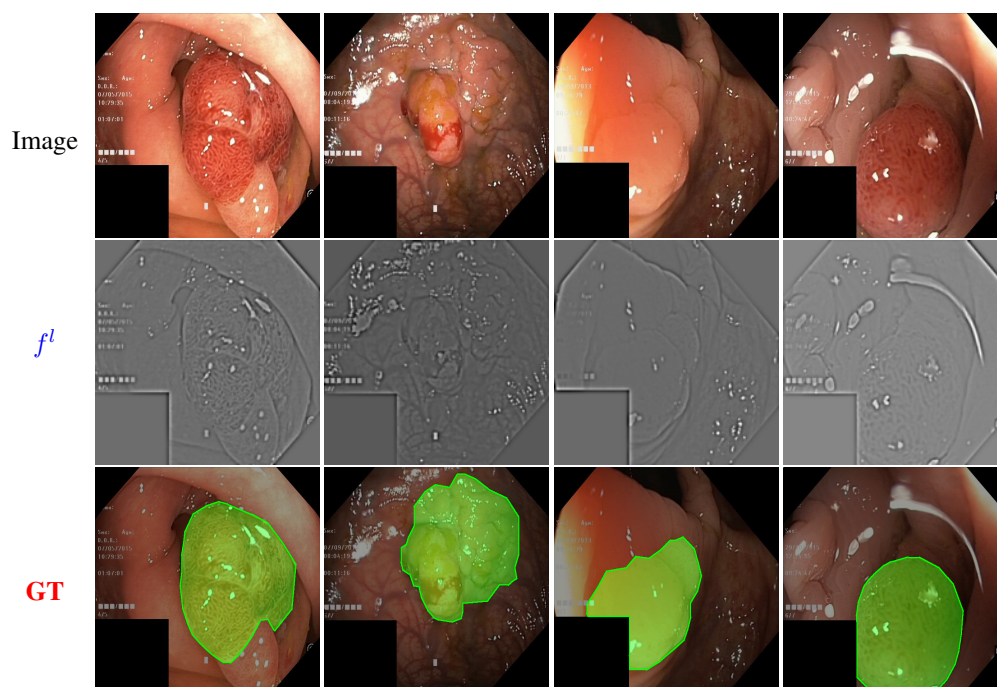


Figure 2. Visualization of the high-frequency feature f^l in Kvasir images.

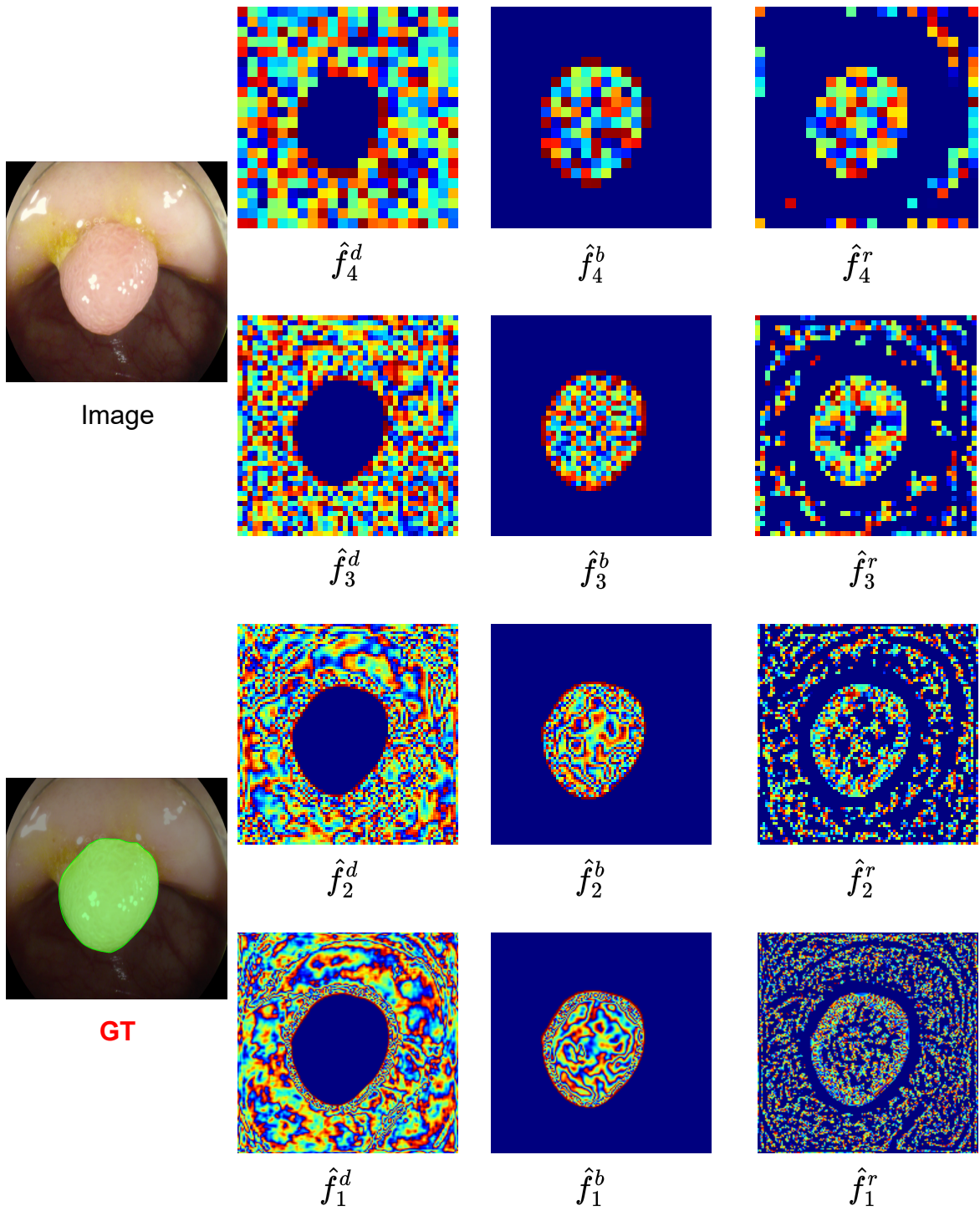


Figure 3. Heatmap visualization of the decoded predicted feature \hat{f}_i^d , the boundary attention map \hat{f}_i^b and the reverse attention map \hat{f}_i^r .