F2Net: A Frequency-Fused Network for Ultra-High Resolution Remote Sensing Segmentation

Supplementary Material

6. Contextual Reasoning

To quantitatively substantiate the semantic reasoning capabilities of F2Net discussed in the main text, we present the class-wise performance gains on the DeepGlobe dataset in Table 8. As shown, local-only classes (e.g., Urban, Water) exhibit moderate improvements. In contrast, context-dependent classes that require global reasoning to resolve ambiguities over large contiguous regions experience substantial gains (e.g., +14.2% for Agriculture and +12.9% for Forest). Furthermore, the model achieves a 94.9% IoU on the Forest class, demonstrating its ability to accurately resolve structural continuity. These class-wise improvements confirm that F2Net's performance gains are largely driven by its enhanced global semantic reasoning rather than merely an increase in generic model capacity.

Method	Urb.	Agr.	Rgl.	For.	Wat.	Bar.	mIoU
Short-Range	78.5	82.0	32.0	82.0	85.5	68.0	71.30
Full F2Net	**81.8**	**96.2**	**34.0**	**94.9**	**87.6**	**86.8**	**80.22**
Gain (Δ)	*+3.3*	***+14.2***	*+2.0*	***+12.9***	*+2.1*	***+18.8***	*+8.92*

Table 8. Class-wise mIoU performance on DeepGlobe.

7. Representation and Stability

Feature Distribution & HFF. The internal downsampling in the low-frequency branch is a strategic design intended to trade spatial resolution for channel capacity, efficiently encoding global texture patterns into feature channels. We validated this by measuring the Average Gradient Magnitude (AGM) Retention Ratio across 30 sampled UHR images. Our empirical results demonstrate that the high-frequency branch maintains significant structural encoding (retaining 83% at Stage 1 and 42% at Stage 4). Conversely, the low-frequency branch acts effectively as a low-pass filter, with the AGM retention dropping sharply to 9% at its final stage. The low cosine similarity (0.22) between the final representations of the two branches further confirms that they capture unique and complementary information.

Moreover, a significant distribution gap exists between the branches: high-frequency features are sparse (Laplacian-like), while low-frequency features are dense (Gaussian-like). Our Hybrid-Frequency Fusion (HFF) module is explicitly designed to project these heterogeneous features into a unified subspace for alignment via a dense interaction matrix. Replacing HFF with a simple linear aggregation module (e.g., RAF) causes a 1.3% mIoU drop, as linear aggregation allows dense signals to overwhelm sparse high-frequency details.

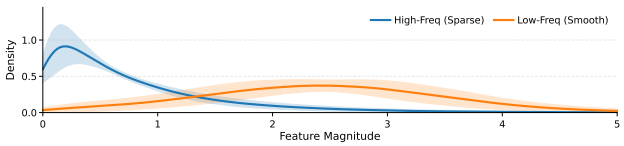

Figure 4. Feature distribution gap between frequency branches. High-frequency features exhibit a sparse, Laplacian-like distribution, while low-frequency features are dense and Gaussian-like.

Gradient Stability. The instability often observed in frequency-aware multi-branch learning is theoretically grounded in the structural imbalance of gradient magnitudes across branches during backpropagation. To investigate this, we tracked the gradient norm ratios across branches during training. As illustrated in Figure 5, without the proposed Cross-Frequency Balance Loss (CFBL), the short-range sub-branch dominates the optimization process (Gradient Ratio $\gg 1$). The introduction of CFBL successfully regularizes this ratio to approximately 1, ensuring stable and balanced learning dynamics across all branches.

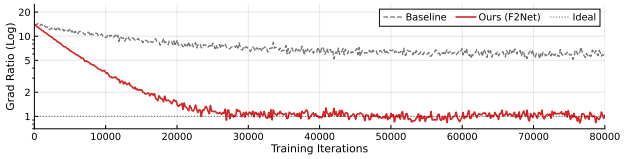

Figure 5. Gradient norm ratios during training. The Cross-Frequency Balance Loss (CFBL) effectively regularizes the ratio to approximately 1, ensuring balanced optimization.