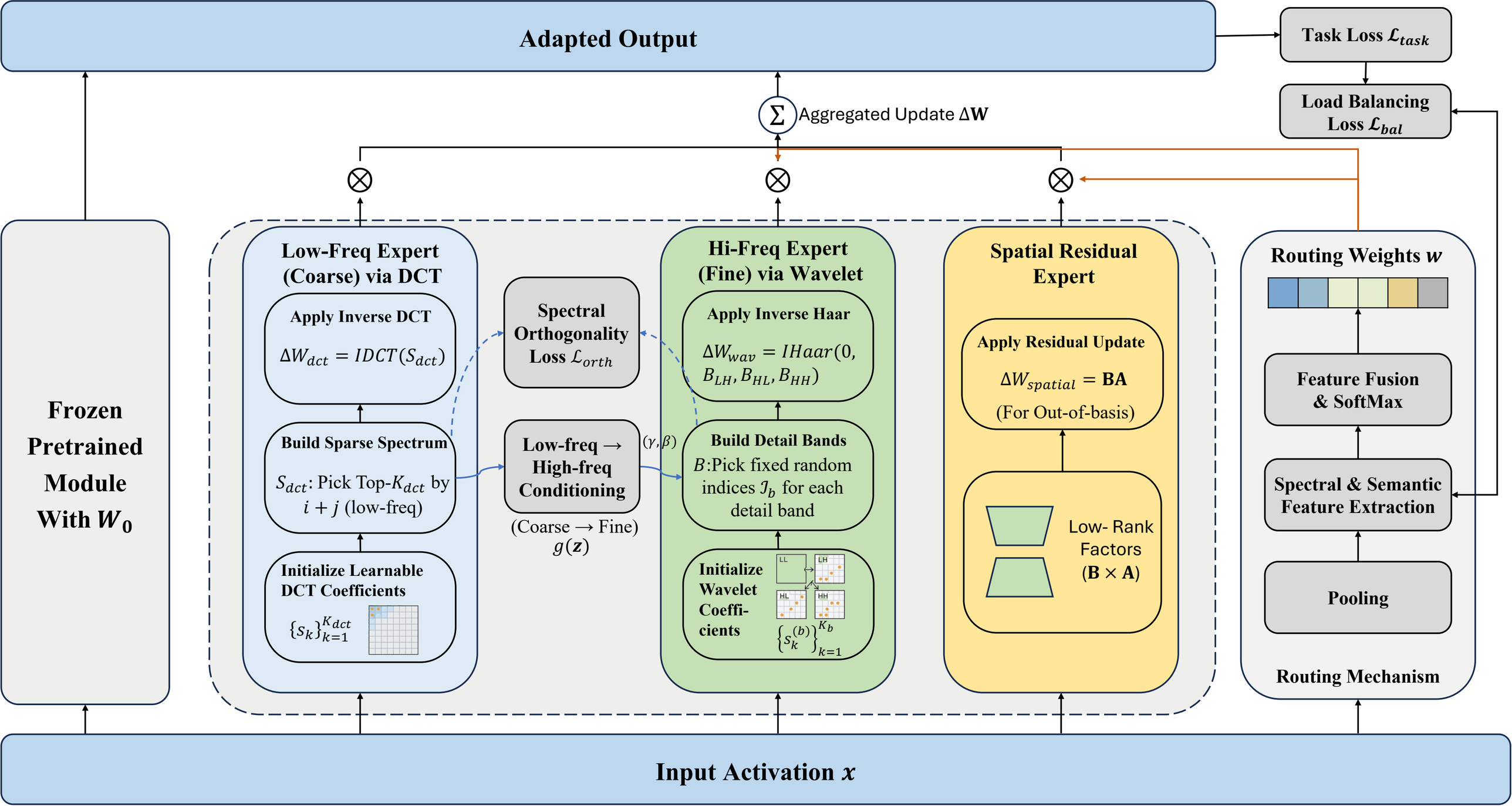
Adapted Output
Task Loss $\mathcal{L}_{task}$
Load Balancing Loss $\mathcal{L}_{bal}$
Aggregated Update $\Delta \mathbf{W}$
Frozen Pretrained Module With W_0
Low-Freq Expert (Coarse) via DCT
Apply Inverse DCT
$\Delta W_{dct} = IDCT(S_{dct})$
Build Sparse Spectrum
S_{dct}: Pick Top-K_{dct} by $i + j$ (low-freq)
Initialize Learnable DCT Coefficients
$\{s_k\}_{k=1}^{K_{dct}}$
Spectral Orthogonality Loss $\mathcal{L}_{orth}$
Low-freq → High-freq Conditioning
(Coarse → Fine) $g(\mathbf{z})$
(γ, β)
Hi-Freq Expert (Fine) via Wavelet
Apply Inverse Haar
$\Delta W_{wav} = IHaar(0, B_{LH}, B_{HL}, B_{HH})$
Build Detail Bands
B:Pick fixed random indices $\mathcal{I}_b$ for each detail band
Initialize Wavelet Coeffi-cients
LL
LH
HL
HH
$\{s_k^{(b)}\}_{k=1}^{K_b}$
Spatial Residual Expert
Apply Residual Update
$\Delta W_{spatial} = \mathbf{BA}$
(For Out-of-basis)
Low- Rank Factors ($\mathbf{B} \times \mathbf{A}$)
Routing Weights w
Feature Fusion & SoftMax
Spectral & Semantic Feature Extraction
Pooling
Routing Mechanism
Input Activation x