Justification for Architectural Modification

The improved model refines the baseline Attention U-Net architecture, introducing several modifications to address class imbalance, improve spatial focus, and stabilise training on deforestation segmentation tasks. The improved model explicitly supports four-channel input (RGB + NIR), utilising Sentinel-2's near-infrared band, which is critical for differentiating vegetation. The baseline model is configured for three-channel RGB input, limiting its ability to separate forest from non-forest areas in dense tropical environments.

Loss Function

The baseline model uses standard binary cross-entropy loss with accuracy and mean squared error (MSE) metrics, which treats all pixels equally and is poorly suited for highly imbalanced segmentation problems where forest pixels dominate. In contrast, the improved model combines weighted binary cross-entropy with focal Tversky loss. The class imbalance was directly addressed by the weighted Binary Cross-Entropy (BCE), which increases the penalty for misclassifying deforested pixels. While focal Tversky loss highlights hard-to-classify boundary regions and reduced false negatives, this combination improves the model's sensitivity to small and fragmented deforestation patches, which are common in the Cameroon context.

Metric selection

The baseline model relies on accuracy and mean squared error, which are not informative for pixel-wise segmentation with skewed class distributions. The improved model replaces these with Intersection-over-Union (IoU) and F1-score, which better capture spatial overlap and balance precision–recall trade-offs. These metrics are also used for model checkpointing, ensuring that saved models reflect segmentation quality rather than raw pixel accuracy.

Regularisation and training stability

Batch normalisation and dropout were used for improved regularisation and training stability within the convolutional blocks. While the baseline includes dropout inconsistently and omits batch normalisation, the enhanced architecture normalises feature distributions after each convolution and applies dropout uniformly. This reduces internal covariate shift, improves gradient flow, and limits overfitting on a relatively small dataset.

Attention gate implementation

The improved attention gates dynamically resample gating signals using bilinear upsampling to ensure spatial alignment between encoder and decoder features. This avoids hard-coded pooling assumptions present in the baseline attention block and improves numerical stability when handling variable feature-map sizes.

Hyperparameter Tunning process

Hyperparameter tuning was performed iteratively based on validation IoU and F1-score rather than loss alone. The learning rate was increased from 1e-4 in the baseline to 5e-4 after observing slow convergence and early stagnation in validation metrics. This higher learning rate accelerated feature learning without causing instability because of batch normalisation. The dropout rate was reduced from 0.25 in the baseline to 0.2, balancing regularisation with the need to preserve fine spatial details such as narrow deforestation boundaries. Higher dropout values suppressed small-scale features critical for accurate segmentation. The filter base size was fixed at 16, maintaining comparable model capacity to the baseline while avoiding excessive parameter growth that could lead to overfitting, given the limited dataset size. For the loss function, the positive class weight in weighted BCE was set to 3.0 after testing showed that lower values under-penalised deforestation pixels. In comparison, higher values led to unstable gradients. The Tversky parameters (α = 0.7, β = 0.3) were chosen to prioritise recall over precision, showing the higher cost of missing deforestation events compared to minor over-segmentation. The focal parameter γ = 0.75 further concentrated learning on difficult boundary pixels.

The hyperparameter choices were tuned based on the characteristics of the Cameroon dataset, which contained class imbalance, small deforested regions, and noisy boundaries. This resulted in improved convergence, higher IoU, and better detection of fragmented forest loss compared to the baseline model.