V2 Experimental Conclusion (Detailed Explanation)

In Version 2 of the model, class-weighted Dice loss combined with categorical cross-entropy loss was introduced to address the severe class imbalance present in the dataset. The primary motivation behind this modification was to ensure that minority classes, which occupy significantly fewer pixels compared to dominant classes such as sky and landscape, contributed more meaningfully to the training process.

This change successfully stabilized the training dynamics. The loss curves became smoother, convergence was more consistent, and the model avoided collapsing into trivial solutions dominated by majority classes. Additionally, predictions began to visibly include minority classes such as bushes, logs, and clutter, indicating that the weighting strategy improved class sensitivity during optimization.

However, despite these improvements, the Mean Intersection-over-Union (IoU) did not improve compared to the baseline model. This observation highlights an important limitation: while the loss function influenced how much the model cared about each class, it did not improve how accurately the model localized those classes in space. IoU is highly sensitive to boundary precision and pixel-level alignment, and the predictions remained spatially coarse, with blurred boundaries and class bleeding.

These results indicate that the core limitation lies not in the loss formulation but in the model's spatial feature representation. The UNet architecture lacks mechanisms to selectively focus on small or sparse regions, causing fine-grained structures to be overwhelmed by dominant background features. Consequently, further improvements require architectural enhancements—such as attention mechanisms or multi-scale feature aggregation—to guide the model toward better spatial focus and improved segmentation quality.