Figure 4. Spatial Entropy Distribution Across CLIP Layers on Multiple Datasets. Visualization showing spatial entropy values across CLIP layers for (a) ADE20K, (b) Cityscapes, (c) COCO-Stuff, and (d) Pascal Context datasets. Darker regions indicate lower spatial entropy. The spatial entropy computation methodology is described in Figure 6.
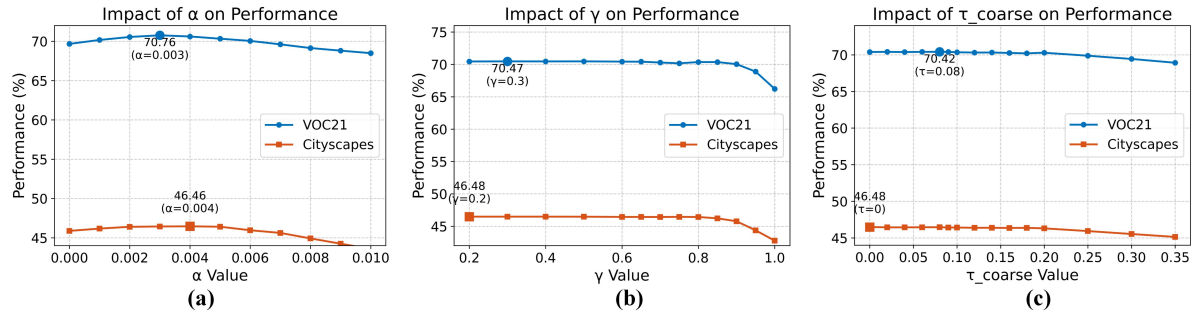

Figure 5. Overall architecture of the proposed S2CLIP. The framework consists of hierarchical progressive feature extraction from CLIP, feature fusion between CLIP and DINO, and SAM post-processing stage.