

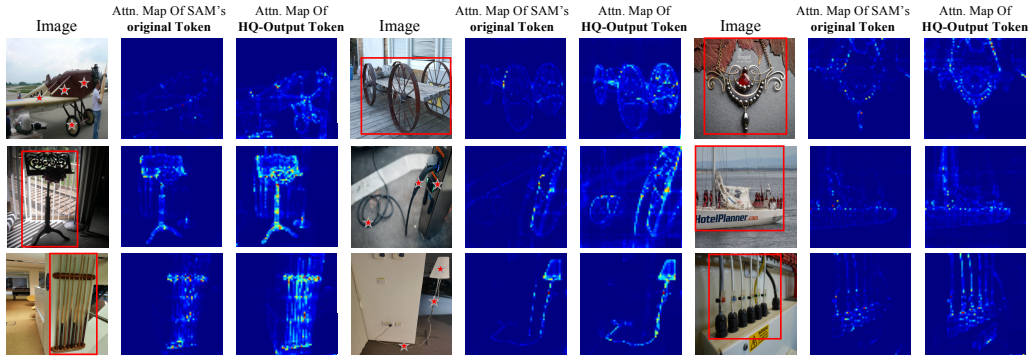


Figure 1: Cross-attention of SAM's original token vs. HQ-Output Token in the last decoder layer. HQ-Token attends to the boundary and thin structure regions that are missed by the original token.

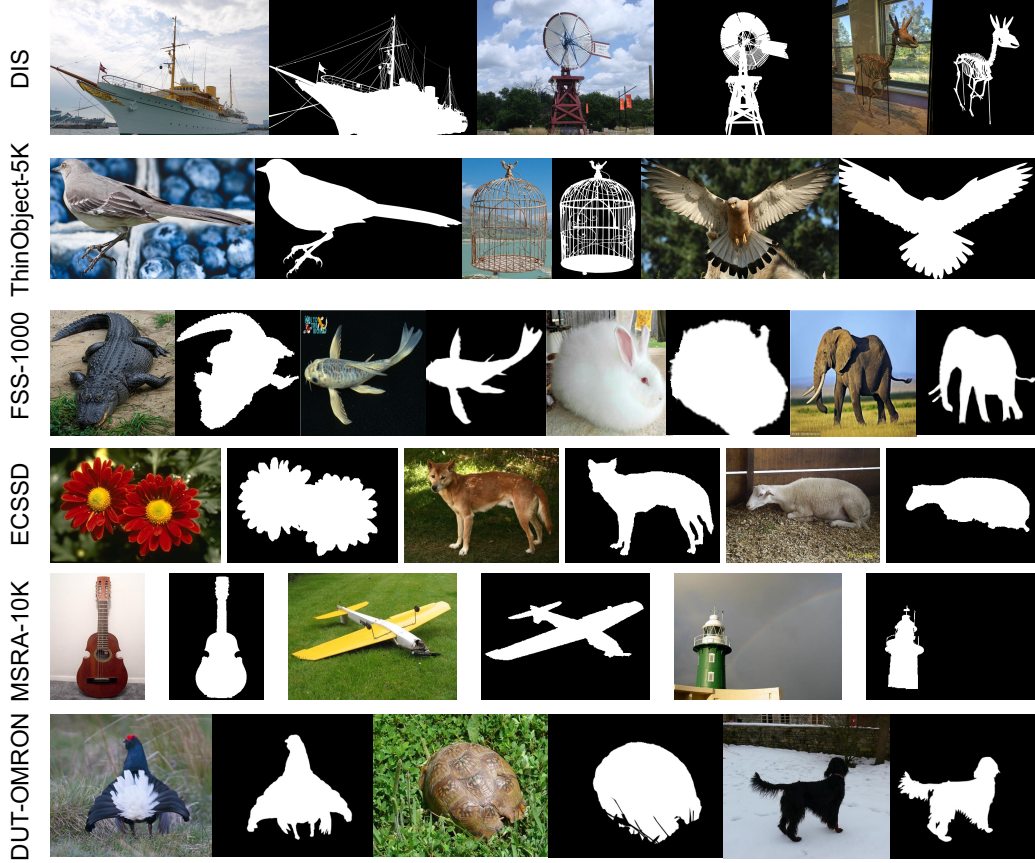


Figure 2: Visualization of annotated mask quality for randomly selected cases from the six dataset components of the HQ-Seg-44K. Zoom in for better viewing the fine-grained mask details.

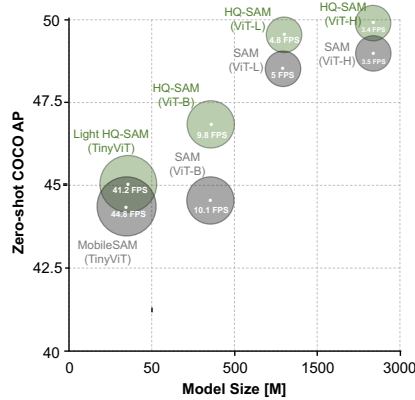


Figure 3: Performance vs. speed vs. model size for an array of SAM variants.