

Referring object detection and referring image segmentation are important tasks that require joint understanding of visual information and natural language. Yet