

Referring expression grounding aims at locating certain objects or persons in an image with a referring expression, where the key challenge is to comprehend a