The task in referring expression comprehension is to localize the object instance in an image described by a referring expression phrased in natural language. A