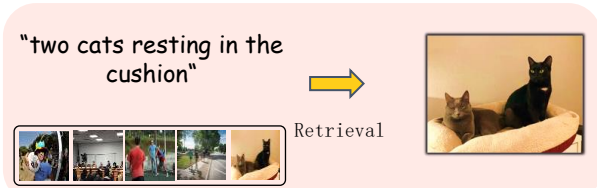
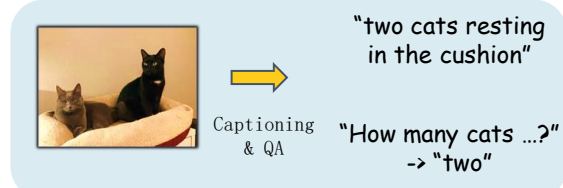


Association:
Image Retrieval

Instance-level Association:
Textual Grounding on Images/Videos
(also known as referring expression comprehension)



Interpretation & Understanding:
Image Caption, VQA