

Visual Question Answering (VQA) requires a fine-grained and simultaneous understanding of both the visual content of images and the textual content of questions.