Visual question answering (VQA) demands simultaneous comprehension of both the image visual content and natural language questions. In some cases, the r