

Visual Question Answering (VQA) in its ideal form lets us study reasoning in the joint space of vision and language and serves as a proxy for the AI task of scene