

Visual Reasoning remains a challenging task, as it has to deal with long-range and multi-step object relationships in the scene. We present a new model for Vis