

The image, question (combined with the history for de-referencing), and the corresponding answer are three vital components of visual dialog. Classical visual d