In this paper, we propose a novel end-to-end trainable Video Question Answering (VideoQA) framework with three major components: 1) a new heterogeneous