KnowIT VQA: Answering Knowledge-Based Questions about Videos

Current VQA limitations:
1) Temporal coherence → VideoQA
2) External Knowledge → KBVQA

So far, they are independent problems.

We address VideoQA and KBVQA together.

Motivation

KnowIT VQA Dataset

• Human expert annotators.
• 207 episodes, 9 seasons.
• 12,087 video clips.
• 24,282 questions pairs.

For each question:
• 4 candidate answers.
• Grounded knowledge.
• Knowledge type (recurrent, specific)

Current VQA limitations:
1) Temporal coherence → VideoQA
2) External Knowledge → KBVQA

We address VideoQA and KBVQA together.

Results and Examples

Visual features:
- ResNet,
- Facial,
- Concepts,
- Captions.

BERT scoring input
\[[CLS] + q_i + a_i 0 + a_i 1 + a_i 2 + [SEP] + k g_j + [SEP]\]

BERT reasoning input
\[[CLS] (+ caps) + subs + q + [SEP] + a^c + kg + [SEP]\]