CRCV REU 2019

Week 1

Literature Review

- TVQA: Localized, Compositional Video Question Answering
  - EMNLP 2018, J. Lei
- Compositional Attention Networks for Machine Reasoning
  - ICLR 2018, D. Hudson
- Natural Language Processing Methods
  - Continuous Bag of Words
  - Skipgram
  - Negative Sampling
  - Word2Vec

TVQA Dataset

- 460 hours of video
- 152,545 Question and Answer Pairs
- 21,793 clips (60-90 sec)

- Multimodal Compositionality
  - Video-QA
  - Associated natural language (subtitles)

Questions

- Main Question part
- Grounding part
  - Temporal Localization
- Each clip has 7 questions
- Each question has 5 multiple choice answers

00:00.755 --> 00:02.655
(Chandler:) Go to your room!
00:06.961 --> 00:08.622
(Janice:) I gotta go, I gotta go.

00:08.829 --> 00:10.057
(Janice:) Not without a kiss.
00:10.264 --> 00:12.391
(Chandler:) Maybe I won't kiss you so you'll stay.

00:12.600 --> 00:14.761
(Joey:) Kiss her. Kiss her!
00:16.771 --> 00:19.137
(Janice:) I'll see you later, sweetie. Bye, Joey.

...

00:39.327 --> 00:40.760
(Chandler:) She makes me happy.
00:41.596 --> 00:44.087
(Joey:) Okay. All right.

...

What is Janice holding on to after Chandler sends Joey to his room?

A Chandler's tie
B Chandler's hands
C Her Breakfast
D Her coat
E Chandler's coffee cup.

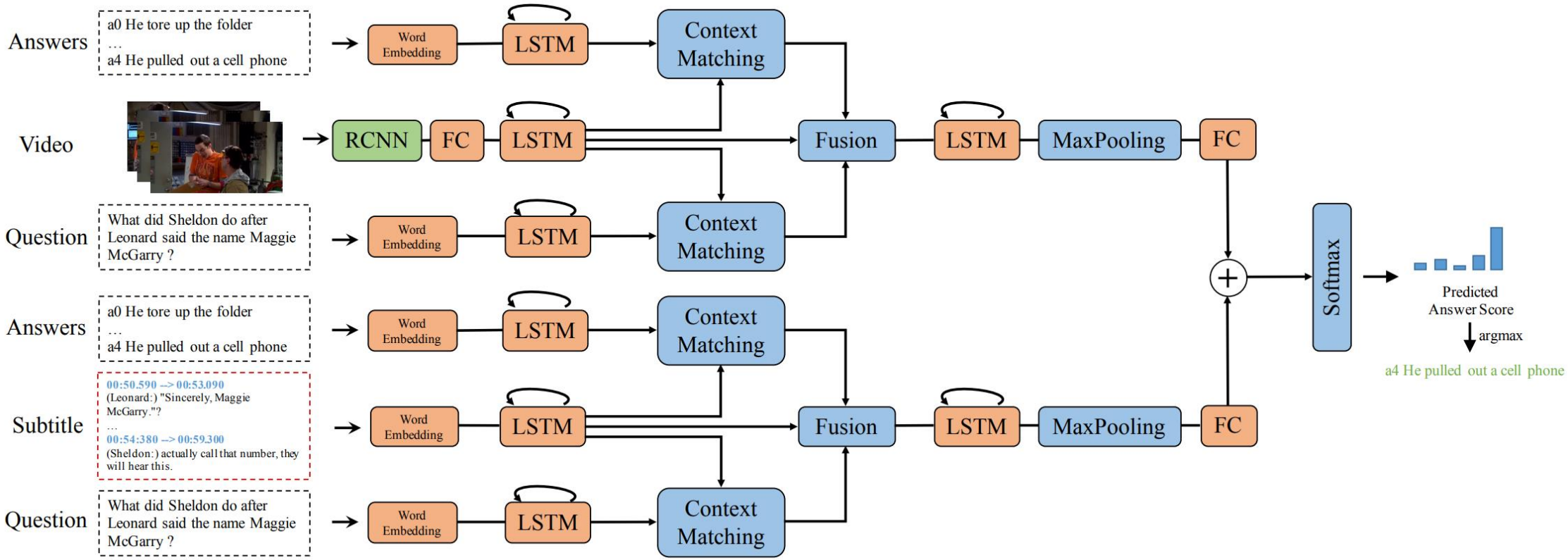
Why does Joey want Chandler to kiss Janice when they are in the kitchen?

A Because Joey is glad that Chandler is happy
B Because Joey likes to watch people kiss
C Because then she will leave
D Because Joey thinks Janice is hot
E Because then Chandler will move away from the toast.

What is on the couch behind Joey when he is at the counter?

A A chick
B A soccer ball
C A duck
D A pillow
E Janice's coat

Model Used


Answers
a0 He tore up the folder
…
a4 He pulled out a cell phone
Video
Question
What did Sheldon do after Leonard said the name Maggie McGarry ?
Answers
a0 He tore up the folder
…
a4 He pulled out a cell phone
Subtitle
00:50.590 --> 00:53.090
(Leonard:) "Sincerely, Maggie McGarry."?
…
00:54:380 --> 00:59.300
(Sheldon:) actually call that number, they will hear this.
Question
What did Sheldon do after Leonard said the name Maggie McGarry ?
Word Embedding
LSTM
Context Matching
RCNN
FC
Fusion
MaxPooling
Softmax
Predicted Answer Score
argmax
a4 He pulled out a cell phone

Results

Model Used	TVQA + S
Accuracy (%) Reported	65.15%
Replication	65.74%

Results

Model Used	TVQA + S	TVQA + V
Accuracy (%) Reported	65.15%	45.03%
Replication	65.74%	45.25%

Results

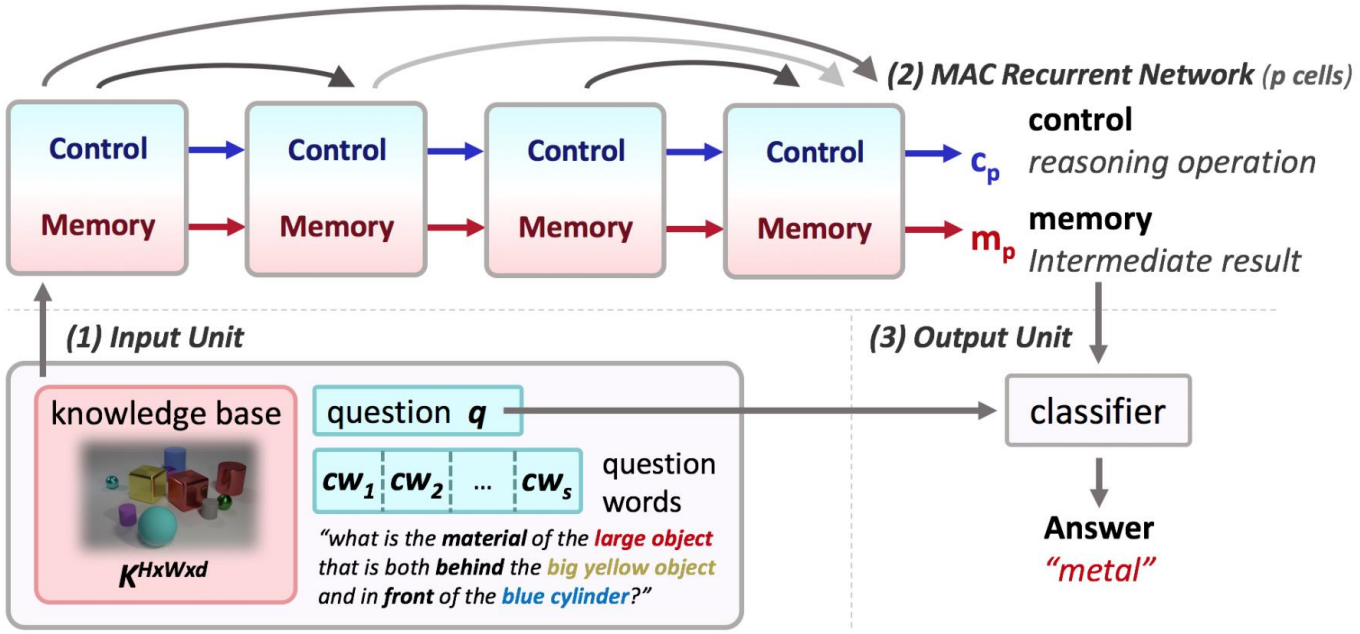
Model Used	TVQA + S	TVQA + V	TVQA + IMG
Accuracy (%) Reported	65.15%	45.03%	43.78%
Replication	65.74%	45.25%	44.42%

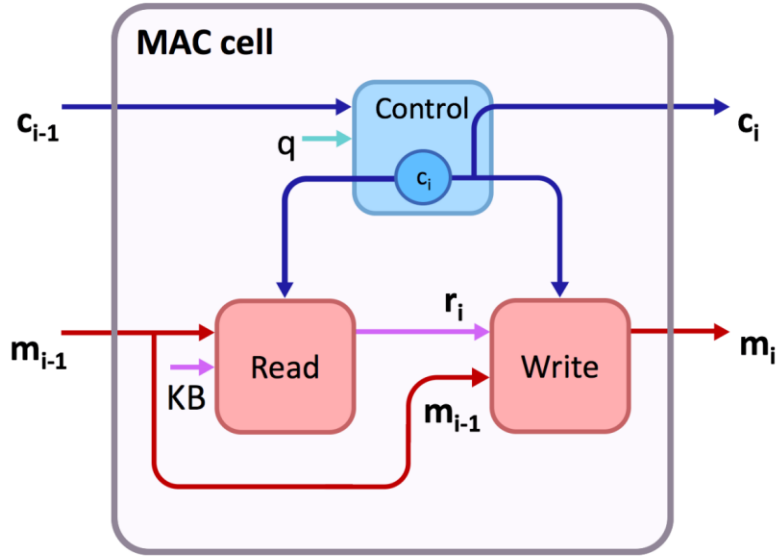
Results

Model Used	TVQA + S	TVQA + V	TVQA + IMG	TVQA + V + IMG
Accuracy (%) Reported	65.15%	45.03%	43.78%	N/A
Replication	65.74%	45.25%	44.42%	45.52%

Compositional Attention Network

- MAC Network
  - Memory, Attention, and Composition
  - Inspired from computer architecture design





Next Steps

- Train different baseline models on the dataset
  - LSTM
  - CNN+LSTM