

MovieQA: Understanding Stories in Movies through Question-Answering

Makarand Tapaswi, Yukun Zhu, Rainer Stiefelwagen,
Antonio Torralba, Raquel Urtasun, Sanja Fidler

Introduction and Dataset

VQA



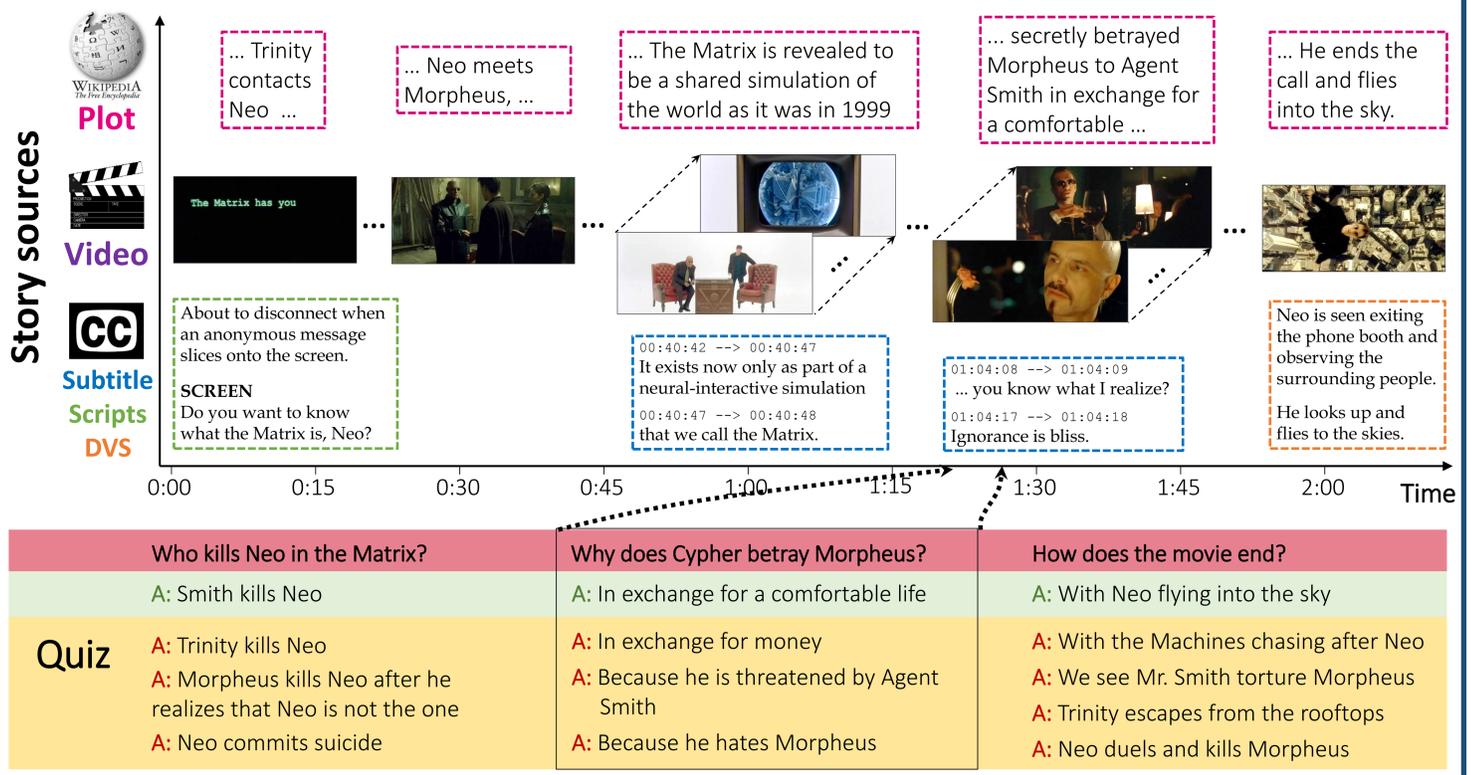
(Malinowski and Fritz 2014), (Antol et al. 2015), (Yu et al. 2015)

- Object semantics: What is ...
- Colors: What color ...
- Yes/No: Does the...
- Counting: How many ...

no reasoning

- does not capture image semantics or story
- essentially classical vision tasks with machine dialog

Story Question-Answering



Benchmark dataset

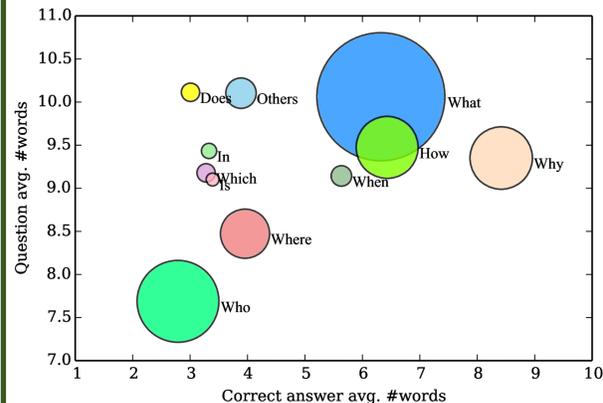
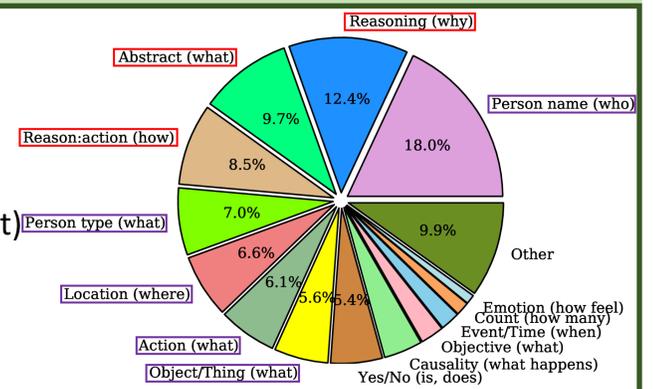
- 408 movies, 14,944 questions, 5 *deceiving* answer options
- video clips with mean duration 200 seconds, 46 shots

	TRAIN	VAL	TEST	TOTAL
Movies with Plots and Subtitles				
# Movies	269	56	83	408
# QA	9,848	1,958	3,138	14,944
Q #words	9.3	9.3	9.5	9.3 ± 3.5
CA. #words	5.7	5.4	5.4	5.6 ± 4.1
WA. #words	5.2	5.0	5.1	5.1 ± 3.9
Movies with Video Clips				
# Movies	93	21	26	140
# QA	4,318	886	1,258	6,462
# Video clips	4,385	1,098	1,288	6,771
Clip dur. (s)	201	198.5	211.4	202.7 ± 216

Analysis of question types

answer semantics

- 30% reasoning (why, how)
- 40% visual cues (who, where, what)



question's first word

- What, Who, Why, How, largest question types
- Why questions have long answers
- Does (yes/no) with reasons

Comparing QA datasets

	Txt	Img	Vid	Goal	Data source	Answering	#Q	AW
MCTest	✓	-	-	reading comprehension	children's stories	MC (4)	2.6K	3.4
bAbI	✓	-	-	reasoning for toy tasks	synthetic	Word	40K	1.0
News article	✓	-	-	information abstraction	news articles	Word	1M*	1*
DAQUAR	-	✓	-	visual: counts, colors, objects, ...	NYU-RGBD	Word List	12.5K	1.2
Visual Madlibs	-	✓	-	visual: scene, objects, people, ...	COCO + Prompts	FITB/MC (4)	150K	2.6
VQA (v1)	-	✓	-	visual understanding	COCO + Abstract	Open/MC (18)	764K	1.2
MovieQA	✓	✓	✓	movie story comprehension	movie stories	MC (5)	15K	5.3

MCTest Richardson et al. MCTest: A Challenge Dataset for the Open-domain Machine Comprehension of Text. In *EMNLP*, 2013.
bAbI Weston et al. Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks. arXiv, 2014.
News Hermann et al. Teaching Machines to Read and Comprehend. In *NIPS*, 2015
DAQUAR Malinowski and Fritz. A Multi-World Approach to Question Answering about Real-World Scenes based on Uncertain Input. In *NIPS*, 2014
VMadlibs Yu et al. Visual Madlibs: Fill in the Blank Image Generation and Question Answering. In *ICCV*, 2015
VQA Antol et al. VQA: Visual Question Answering. In *ICCV*, 2015

More examples

Q. What does Morpheus offer Neo when they meet?
A1. A yellow pill and green pill
A2. A red pill and blue pill
A3. Vitamins



Q. What is Nebuchadnezzar?
A1. A restaurant
A2. Agent Smith's hovercraft
A3. Morpheus's hovercraft

SCRIPT
MORPHEUS: Come with me, see for yourself.
MORPHEUS: This is my ship, the Nebuchadnezzar. It's a hovercraft.

Q. Does Morpheus escape the Agents?
A1. Yes, Neo and Trinity help
A2. Yes, Morpheus drugs them
A3. No, he is captured and tortured

PLOT
Neo resolves to return to the Matrix to rescue Morpheus, and Trinity accompanies him. They succeed in rescuing Morpheus.