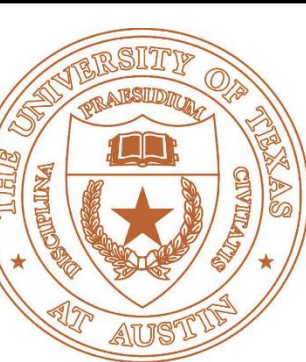


Let's Agree to Disagree: A Meta-Analysis of Disagreement Among Crowdworkers During Visual Question Answering

Anuparna Banerjee, Samridhi Ojha and Danna Gurari



Visual Question Answering (VQA)



Who threw the frisbee?

1. man on right
2. man in green shirt
3. man in green shirt
4. man on right
5. man in green vest
6. man in green shirt
7. man in neon shirt
8. man with green shirt
9. cannot tell
10. man on right



Who looks happier?

1. Old person
2. man
3. man
4. man
5. Old person
6. man
7. man
8. man
9. man
10. grandpa



Are the bikes safe?

1. yes
2. yes
3. yes
4. no
5. yes
6. yes
7. yes
8. no
9. yes
10. yes

Status Quo: Train machines to predict a single answer.

Problem: Crowd disagrees over the answers to a visual question.

Datasets

VQA Real Images
48 Visual Questions



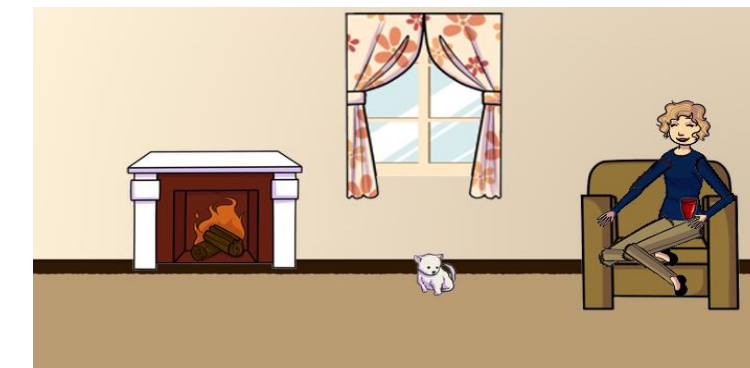
How many chairs are there in this picture?

VizWiz Images
48 Visual Questions



What item is this?

VQA Abstract Scenes
48 Visual Questions



What color are the plates?

Entropy of Answers



What is the woman doing?

1. Smiling
2. watching
3. watching boy
4. standing
5. laughing
6. watching boy
7. standing
8. laughing
9. standing
10. watching



What color is the carpet?

1. beige
2. beige
3. beige
4. brown
5. tan
6. beige
7. tan
8. tan
9. brown
10. cream



Does the woman have curly hair?

1. yes
2. yes
3. yes
4. yes
5. yes
6. yes
7. no
8. yes
9. no
10. yes

High Entropy Answers

Low Entropy Answers

User Interface: Radio Button

Question: What color is the man's jacket?

Answer Choices:

1. beige
2. tan
3. beige
4. tan
5. khaki
6. tan
7. tan
8. tan
9. tan
10. tan



- ☐ Task Difficulty ☐ Ambiguity ☐ Granularity
☐ Subjectivity ☐ Insufficient Visual Evidence ☐ Synonym

User Interface: Checkbox

Question: What does this can say?

Answer Choices:

1. unanswerable
2. Unanswerable
3. Bag of Hall's Cough Drops
4. unanswerable
5. Unanswerable
6. unanswerable
7. Halls
8. unsuitable image
9. STORE ROOM
10. unsuitable image



- ☐ Task Difficulty ☐ Ambiguity ☐ Granularity
☐ Subjectivity ☐ Insufficient Visual Evidence ☐ Synonym

Reasons of disagreement [1]

- Task Difficulty
- Granularity
- Insufficient Visual Evidence (IVE)
- Ambiguity
- Subjectivity
- Synonym

References

[1] D. Gurari and K. Grauman. "CrowdVerge: Predicting If People Will Agree on the Answer to a Visual Question." ACM Conference on Human Factors in Computing Systems (CHI), Denver, CO, May, 2017.

Results: Radio Button Interface

	0.5 to 1.0	1.0 to 1.5	more than 1.5
VQA Real Images	Subjectivity	Subjectivity	Subjectivity
VQA Abstract Images	Subjectivity	Subjectivity	Subjectivity
VizBiz Images	IVE	Synonym; IVE	Synonym

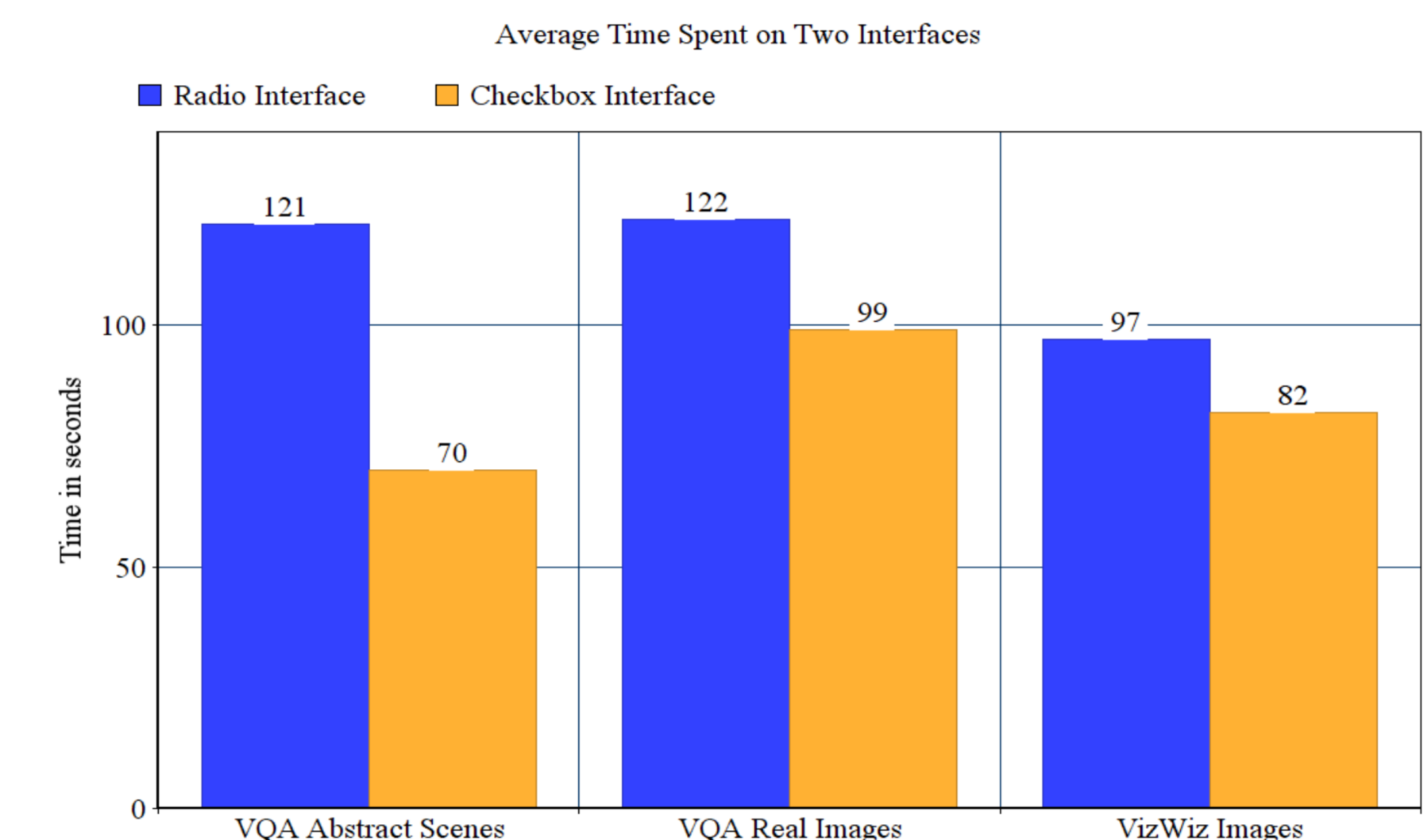
Subjectivity is the major reason of disagreement for *VQA Real* and *Abstract Images*.

Insufficient Visual Evidence (IVE) is the major reason of disagreement for *VizWiz Images*.

Results: Checkbox Interface

	0.5 to 1.0	1.0 to 1.5	more than 1.5
VQA Real Images	Subjectivity	Subjectivity	Subjectivity
VQA Abstract Images	Subjectivity	Subjectivity	Subjectivity
VizBiz Images	IVE	IVE	Synonym

Checkbox interface produces similar results as that of the radio button interface.



The crowd provides results faster on a checkbox interface than a radio button interface.