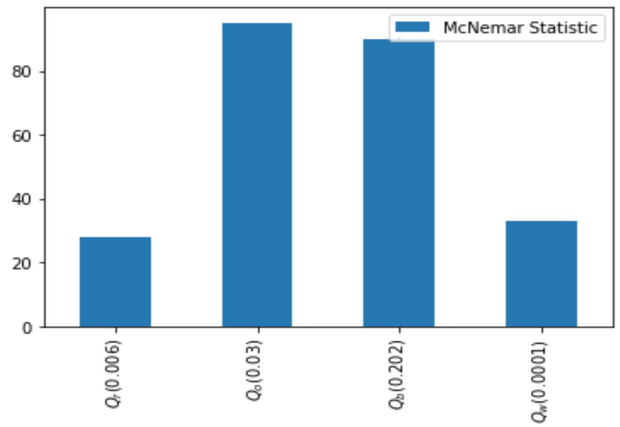
Question Modifiers in Visual Question Answering

William Britton, Somdeb Sarkhel, and Deepak Venugopal
University of Memphis, Adobe Research, University of Memphis



Significance of performance drop for adding modifiers along several axes

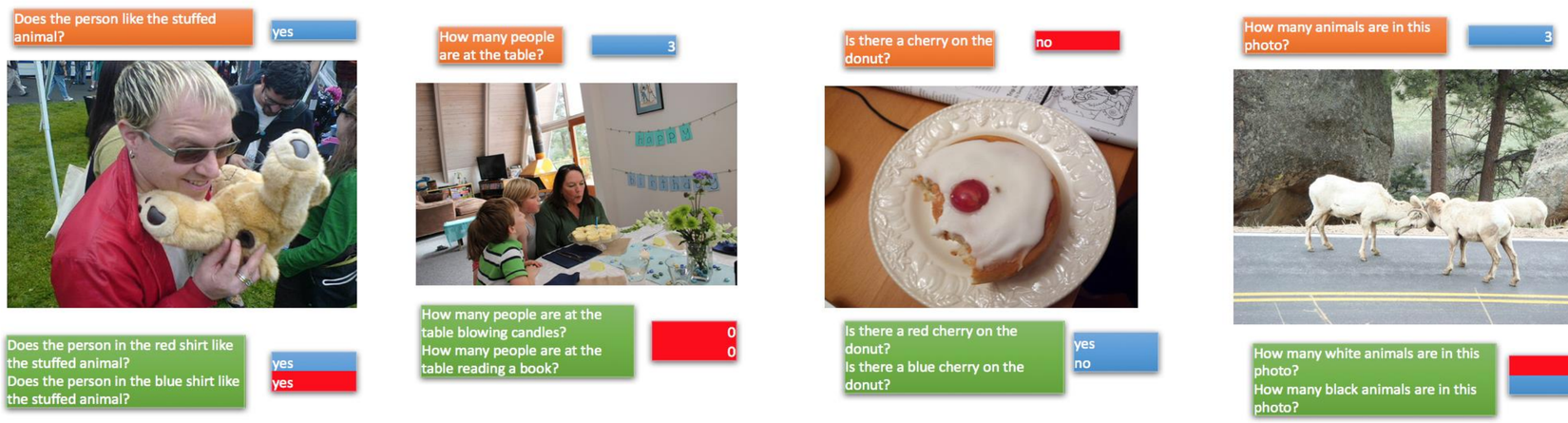
Q_r = relational modifiers
Q_o = object modifiers
Q_b = binary questions
Q_w = non-binary questions

- **Modifiers**
- Object Property – color, material, shape, size, expression, behavior
- Relational – relative spatial relationships (on, in, around, above, below, next to, in front of, behind, etc.)

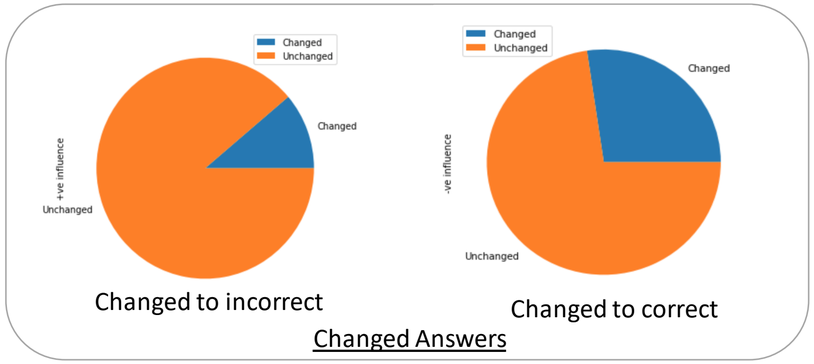
Contribution Evaluating sensitivity of state of the art VQA model (LXMERT) to specific types information in questions such as modifiers to subjects and objects

- To create the dataset, a 3065 question subset of the one-word answer questions in the full VQA2.0 dataset was used
- For every question q, a pair of questions with modifiers were generated by MTurk workers where one preserves the same answer and the other changes the answer
- Each modified question corresponds to one of two types: object property and relational
- LXMERT model trained on modified questions and evaluated performance for modification types

Object property - Observed a significant decrease in performance for modified questions
Relational - Observed a significant decrease in performance for modified questions
Binary questions - No significant decrease in performance for modified questions
Non-binary questions - Significant decrease in performance for modified questions



Example predictions made by the model for (q0, q1, q2). Correct predictions are marked by blue and wrong ones by red.



Changed Answers

Percentages where answers to q_1 were changed from the original answer in q_0

$\|Q\|$	$\|Q_b\|$	$\|Q_w\|$	$\|Q_r\|$	$\|Q_o\|$
3065	1968	1564	1105	2427

Summary of data collected. Each element of Q is a triple (q_0, q_1, q_2) where q_0 is the original question, q_1 is modified question with the same answer, q_2 is modified question with different answer

Q_r	Q_o	Q_b	Q_w
0.497	0.540	0.628	0.398

Evaluation of accuracy for modifiers and answer types

Original (q_0)	Modified (q_1)	Modified (q_2)	$\delta(q_0, q_2)$
0.647	0.61	0.442	0.453

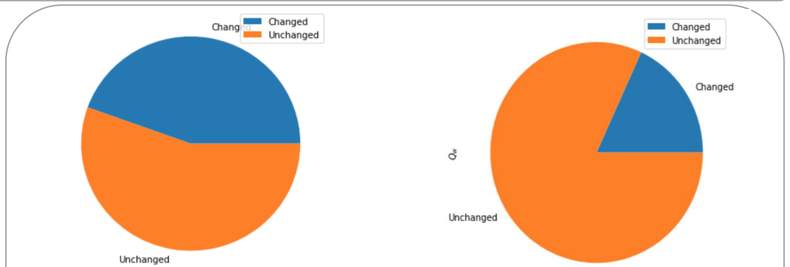
Evaluating influence of question modifiers. Accuracy for LXMERT is computed for the original questions (q_0) and both the answer preserving (q_1) and non answer preserving (q_2) rephrased questions.

	Relational Modifiers	Object Modifiers
Yes/No Answers	0.655	0.697
Open-ended Answers	0.475	0.537

Evaluating the joint influence of modifiers and answer types. The accuracy values for LXMERT are shown for each case.

Example Question:
Where is the child sitting?

More detailed questions:
Where is the child who is holding a bottle sitting?
Where is the child drinking from a bottle?



Changed Answers for "unknowns"

Percentage that given q_1 answered correctly, q_2 was answered differently for determinable answers vs "unknown" answers