Perception Matters: Detecting Perception Failures of VQA Models Using Metamorphic Testing

Supplementary Materials

Yuanyuan Yuan[1] Shuai Wang[1] Mingyue Jiang[2] Tsong Yueh Chen[3]
[1]HKUST, [2]Zhejiang Sci-Tech University, [3]Swinburne University of Technology
{yyuanaq, shuaiw}@cse.ust.hk, mjiang@stu.edu.cn, tychen@swin.edu.au

A. Errors found Using Question-Oriented MRs

This section provides some errors that we found using question-oriented MRs.

Q: # of pillows and hands? A: 2
Q_1: # of pillows ? A_1: 1
Q_2: # of hands ? A_2: 2

Q: # of girls and boards? A: 2
Q_1: # of girls? A_1: 1
Q_2: # of boards? A_2: 0

Q: # of lines and roads? A: 2
Q_1: # of lines? A_1: 3
Q_2: # of roads? A_2: 1

Q: # of bikes and cyclists? A: 4
Q_1: # of bikes? A_1: 3
Q_2: # of cyclists? A_2: 4

Q: # of trees and frisbees? A: 2
Q_1: # of trees? A_1: 5
Q_2: # of frisbees? A_2: 1

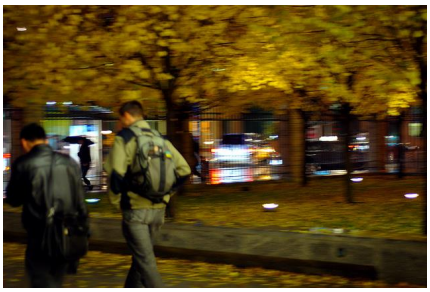
Q: # of horses and trees? A: 2
Q_1: # of horses? A_1: 0
Q_2: # of trees? A_2: 5

Q: # of towels and windows? A: 2
Q_1: # of towels? A_1: 3
Q_2: # of windows? A_2: 2

Q: # of lights and towels? A: 5
Q_1: # of lights? A_1: 3
Q_2: # of towels? A_2: 3

Q: # of televisions and floors? A: 1
Q_1: # of televisions ? A_1: 1
Q_2: # of floors? A_2: 1

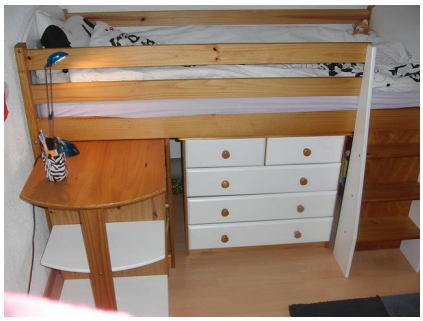
Q: # of blankets and cats? A: 1
Q_1: # of blankets? A_1: 1
Q_2: # of cats? A_2: 2



Q: # of persons and lights? A: 2
Q_1: # of persons ? A_1: 2
Q_2: # of lights? A_2: 2

Q: # of black jackets and green jackets? A: 2
Q_1: # of black jackets ? A_1: 2
Q_2: # of green jackets ? A_2: 2

Figure 1: Errors found on $\text{Oscar}^{+}_{\text{large}}$ using object-/property-oriented partitioning MRs. Due to the limited space, we replace phrase "*what is the total number*" with "#".

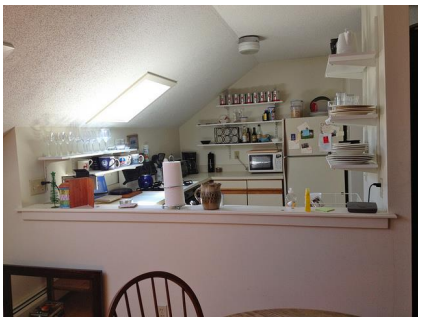


Q: Is there any bed? A: Yes
¬Q: Is there no bed? A: Yes

Q: Is there any yellow bed? A: No
¬ Q: Is there no yellow bed? A: No

Q: Is there any purple dog? A: No
¬Q: Is there no purple dog? A: No

Q: Is there any brick building? A: No
¬Q: Is there no brick building? A: No



Q: Is there any microwave? A: Yes
¬Q: Is there no microwave? A: Yes

Q: Is there any golden chair? A: No
¬Q: Is there no golden chair? A: No

Q: Is there any black spoon? A: Yes
¬Q: Is there no black spoon? A: Yes

Q: Is there any white bowl? A: Yes
¬Q: Is there no white bowl? A: Yes

Q: Is there any sink? A: Yes
¬Q: Is there no sink? A: Yes

Q: Is there any drain? A: Yes
¬Q: Is there no drain? A: Yes

Q: Is there any car? A: Yes
¬Q: Is there no car? A: Yes

Q: Is there any yellow statue? A: No
¬Q: Is there no yellow statue? A: No

Figure 2: Errors found on Oscar$^{+}_{\text{large}}$ using object-/property-oriented reversion MRs. Due to the limited space, we replace phrase "*what is the total number*" with "#".

Q: # of horses and signs? A: 1
Q_3: # of signs and horses? A_3: 2

Q: # of men and trees? A: 2
Q_3: # of trees and men? A_3: 10

Q: # of women and giraffes? A: 2
Q_3: # of giraffes and women? A_3: 4

Q: # of trees and ears? A: 10
Q_3: # of ears and trees? A_3: 2

Q: # of umbrellas and chairs? A: 2
Q_3: # of chairs and umbrellas? A_3: 3

Q: # of kites and persons? A: 2
Q_3: # of persons and kites? A_3: 1

Q: # of watches and bottles? A: 2
Q_3: # of bottles and watches? A_3: 1

Q: # of tables and men? A: 1
Q_3: # of men and tables? A_3: 2

Q: # of dogs and benches? A: 1
Q_3: # of benches and dogs? A_3: 4

Q: # of sidewalks and teeth? A: 2
Q_3: # of teeth and sidewalks? A_3: 0

Q: # of players and pants? A: 2
Q_3: # of pants and players? A_3: 4

Q: # of men and hats? A: 2
Q_3: # of hats and men? A_3: 3

Figure 3: Errors found on $\text{Oscar}^{+}_{\text{large}}$ using object-/property-oriented reordering MRs. Due to the limited space, we replace phrase "*what is the total number*" with "#".

B. Errors found Using Image-Oriented MRs

This section provides some errors that we found using image-oriented MRs.

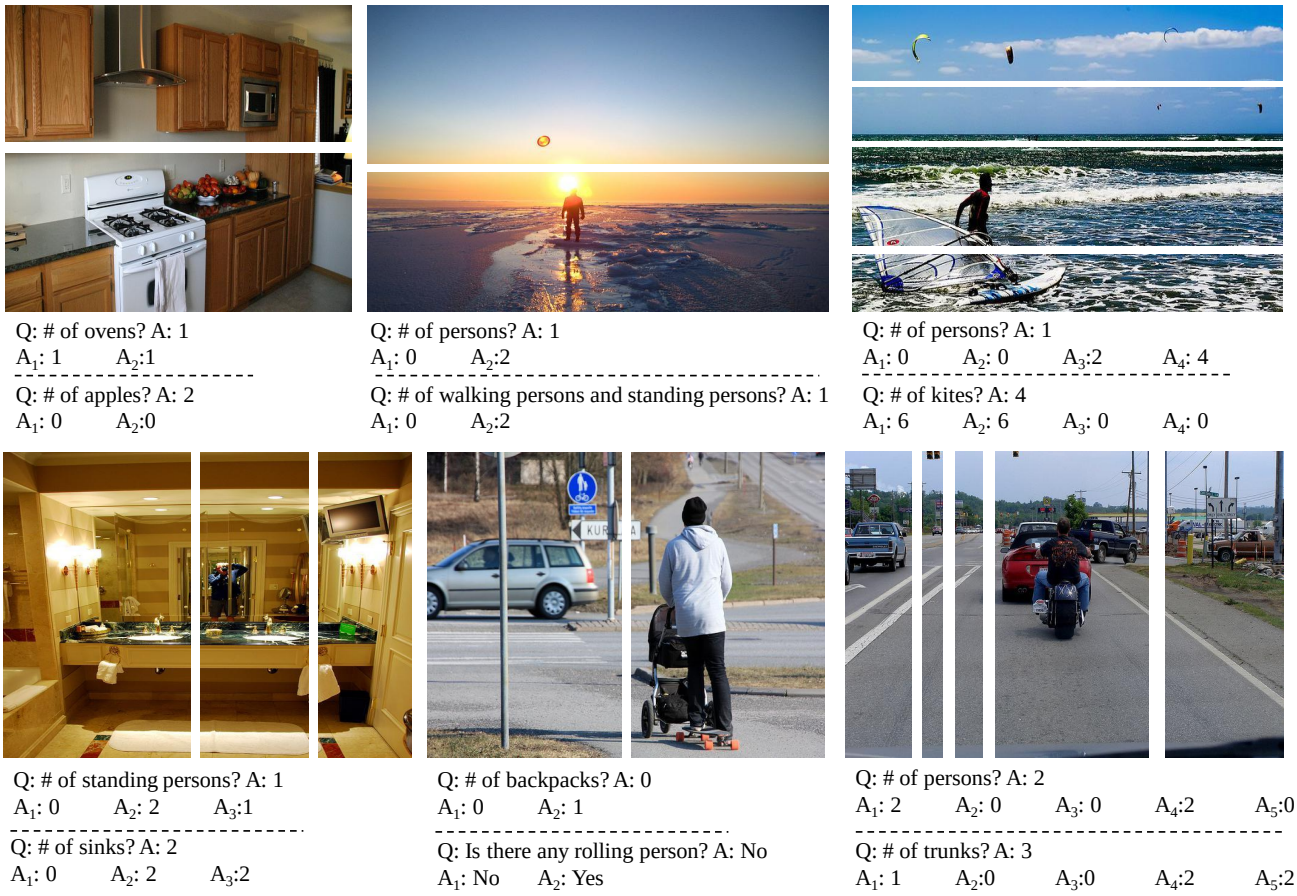


Figure 4: Errors found on $\text{Oscar}^{+}_{\text{large}}$ using image-cutting MRs. Due to the limited space, we replace phrase "*what is the total number*" with "#".

Q: # of blue baseball bats? A: 0
Q: # of brown benches? A: 0

Q: # of red traffic lights? A: 1
Q: # of yellow traffic lights and green traffic lights? A: 1

Q: # of benches? A: 2
Q: Is there no green bench? A: Yes

Q: # blue baseball bats? A: 1
Q: # of brown benches? A: 2

Q: # of red traffic lights? A: 2
Q: # of yellow traffic lights and green traffic lights? A: 0

Q: # of benches? A: 0
Q: Is there no green bench? A: No

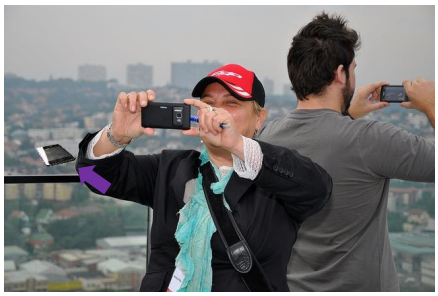
Q: # number of TVs? A: 1
Q: # of mice? A: 2

Q: # of white frisbees? A: 1
Q: # of yellow frisbees? A: 1

Q: # of lying persons? A: 0
Q: # of black handbags: A: 1

Q: # number of TVs? A: 2
Q: # of mice? A: 1

Q: # of white frisbees? A: 2
Q: # of yellow frisbees? A: 0



Q: # of lying persons? A: 2
Q: # of black handbags: A: 2

Figure 5: Errors found on $\text{Oscar}^{+}_{\text{large}}$ using object-insertion MRs. The inserted objects are pinpointed in the figures. Due to the limited space, we replace phrase "*what is the total number*" with "#".

Q: # of gray tennis rackets? A: 0

Q: # of standing persons? A: 2

Q: # of cars and birds? A: 2

Q: # of gray tennis rackets? A: 2

Q: # of standing persons? A: 1

Q: # of cars and birds? A: 1

Q: # of tennis rackets? A: 2

Q: # of standing persons? A: 2

Q: # of walking persons? A: 2

Q: # of tennis rackets? A: 0

Q: # of standing persons? A: 0

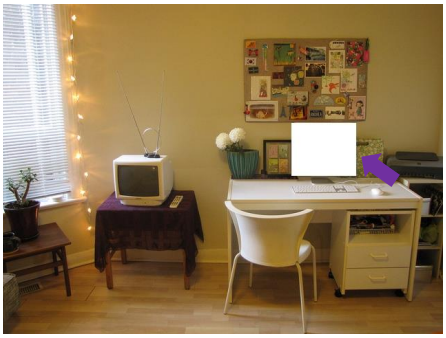
Q: # of walking persons? A: 3

Figure 6: Errors found on Oscar$^{+}_{\text{large}}$ using object-removal MRs. The removed objects are pinpointed in the figures. Due to the limited space, we replace phrase "*what is the total number*" with "#".

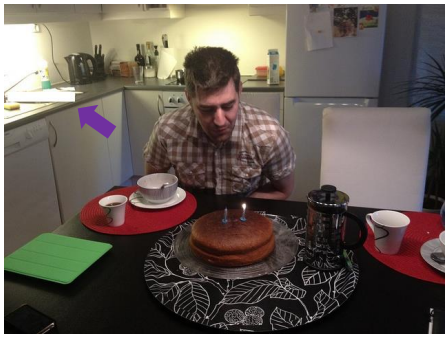
Q: # of potted plants? A: 2

Q: # of potted plants? A: 0

Q: # of golden sinks? A: 3



Q: # of potted plants? A: 1



Q: # of potted plants? A: 1

Q: # of golden sinks? A: 2

Q: # of skating persons? A: 1

Q: # black refrigerators? A: 0

Q: # of blue mice? A: 0

Q: # of skating persons? A: 0

Q: # black refrigerators? A: 1

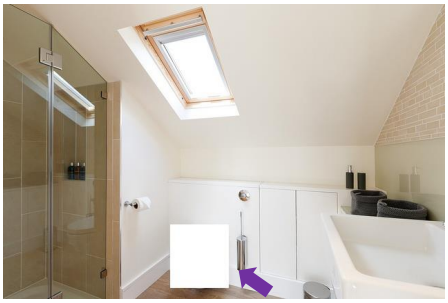
Q: # of blue mice? A: 1

Figure 7: Errors found on Oscar$^{+}_{\text{large}}$ using object-removal MRs. The removed objects are pinpointed in the figures. Due to the limited space, we replace phrase "*what is the total number*" with "#".

Q: # of toilets? A: 1

Q: # of couches? A: 2

Q: # of bottles? A: 2



Q: # of toilets? A: 1

Q: # of couches? A: 2

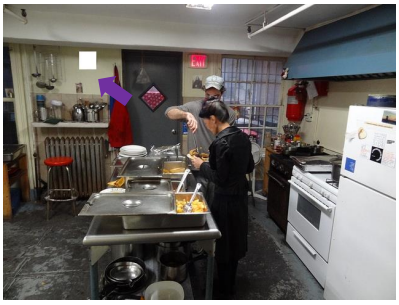
Q: # of bottles? A: 2

Q: # of boats? A: 10

Q: # of clocks? A: 1

Q: # of persons? A: 2

Q: # of boats? A: 4



Q: # of clocks? A: 2

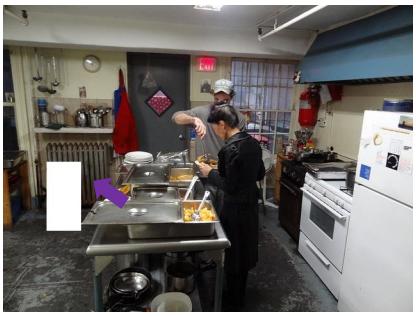
Q: # of persons? A: 2

Figure 8: Errors found on GridFeat+MoViE^{+} using object-removal (**Removal**$^{+}$) MRs. The removed objects are pinpointed in the figures. Due to the limited space, we replace phrase "*what is the total number*" with "#".

Q: # of chairs? A: 2

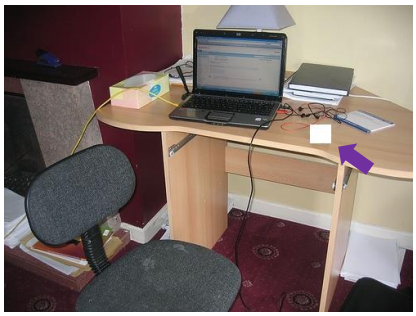
Q: # of cars? A: 2

Q: # of mice? A: 1



Q: # of chairs? A: 2

Q: # of cars? A: 2



Q: # of mice? A: 1

Q: # of persons? A: 3

Q: # of microwaves? A: 1

Q: # of cars? A: 3

Q: # of persons? A: 3

Q: # of microwaves? A: 1

Q: # of cars? A: 3

Figure 9: Errors found on GridFeat+MoViE$^+$ using object-removal (**Removal**$^+$) MRs. The removed objects are pinpointed in the figures. Due to the limited space, we replace phrase "*what is the total number*" with "#".

C. False Positive Cases

This section provides some false positive cases we found on BERT-like models.

Q: What is the total number of cats and couches? A: 0
Q_1: What is the total number of cats? A_1: 0
Q_2: What is the total number of couches? A_2: 0
Q_3: What is the total number of couches and cats? A_3: 0

Q: What is the total number of heads and plants? A: 0
Q_1: What is the total number of heads? A_1: 0
Q_2: What is the total number of plants? A_2: 0
Q_3: What is the total number of plants and heads? A_3: 0

Q: What is the total number of helmets and players? A:0
Q_1: What is the total number of helmets? A_1:0
Q_2: What is the total number of players? A_2:0
Q_3: What is the total number of players and helmets? A_3:0

Q: What is the total number of bats and men? A:0
Q_1: What is the total number of bats? A_1:0
Q_2: What is the total number of men? A_2:0
Q_3: What is the total number of men and bats? A_3:0

Q: What is the total number of birds and dogs? A: 0
Q_1: What is the total number of birds? A_1: 0
Q_2: What is the total number of dogs? A_2: 0
Q_3: What is the total number of dogs and birds? A_3: 0

Q: What is the total number of birds and heads? A: 0
Q_1: What is the total number of birds? A_1: 0
Q_2: What is the total number of heads? A_2: 0
Q_3: What is the total number of heads and birds? A_3: 0

Q: What is the total number of buses and women? A: 0
Q_1: What is the total number of buses? A_1: 0
Q_2: What is the total number of women? A_2: 0
Q_3: What is the total number of women and buses? A_3: 0

Q: What is the total number of men and buildings? A: 0
Q_1: What is the total number of men? A_1: 0
Q_2: What is the total number of buildings? A_2: 0
Q_3: What is the total number of buildings and men? A_3: 0

Figure 10: False positve cases found on VisualBERT.

D. Definition of P_{human} and P_{color}

This section provides the definition of P_{human} and P_{color}. We collect P_{human} and P_{color} from the attributes of Visual G.

P_{color}	P_{human}
black, blue, brown, dark, golden, gray, green, orange, pink, purple, red, white, yellow	throwing, blowing, cooking, flying, skiing, surfing, sleeping, jumping, swimming, crouching, waving, rolling, grazing, walking, smiling, laying, batting, sitting, running, moving, eating, skating, playing, standing, driving, watching, parking, bending, hanging, squatting, riding, landing, resting, looking, holding, racing, kneeling, sliding, lying, serving, hitting, pointing, posing, wearing, swinging, laughing, talking, skateboarding

Figure 11: Definition of P_{human} and P_{color}, which are used in the property-oriented question partitioning scheme.