
Supplementary Materials for “Exploring Models and Data for Image Question Answering”

Mengye Ren¹, Ryan Kiros¹, Richard S. Zemel^{1,2}
 University of Toronto¹
 Canadian Institute for Advanced Research²
 {mren, rkiros, zemel}@cs.toronto.edu

A Supplementary Material

A.1 Question Generation: Syntax Tree Example

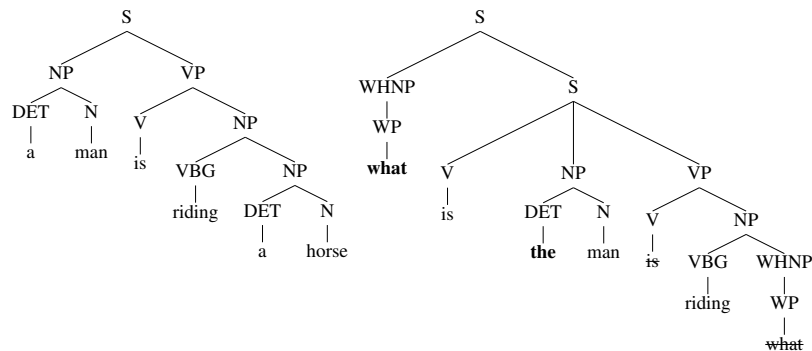


Figure 1: Example: “A man is riding a **horse**” => “**What** is **the** man riding?”

A.2 Post-Processing of COCO-QA Detail

First, answers that appear less than a frequency threshold are discarded. Second, we enroll a QA pair one at a time. The probability of enrolling the next QA pair (q, a) is:

$$p(q, a) = \begin{cases} 1 & \text{if } \text{count}(a) \leq K \\ \exp\left(-\frac{\text{count}(a)-K}{K_2}\right) & \text{otherwise} \end{cases} \quad (1)$$

where $\text{count}(a)$ denotes the current number of enrolled QA pairs that have a as the ground truth answer, and K, K_2 are some constants with $K \leq K_2$. In the COCO-QA generation we chose $K = 100$ and $K_2 = 200$.

A.3 More Sample Questions and Responses

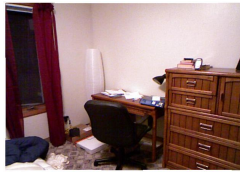


COCOQA 23419
What is the black and white cat wearing?

Ground truth: hat
 IMG+BOW: hat (0.50)
 2-VIS+BLSTM: tie (0.34)
 BOW: tie (0.60)

COCOQA 23419a
What is wearing a hat?

Ground truth: cat
 IMG+BOW: cat (0.94)
 2-VIS+BLSTM: cat (0.90)
 BOW: dog (0.42)



DAQUAR 2136
What is right of table?

Ground truth: shelves
 IMG+BOW: shelves (0.33)
 2-VIS+BLSTM: shelves (0.28)
 LSTM: shelves (0.20)

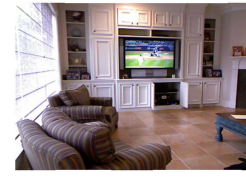
DAQUAR 2136a
What is in front of table?

Ground truth: chair
 IMG+BOW: chair (0.64)
 2-VIS+BLSTM: chair (0.31)
 LSTM: chair (0.37)



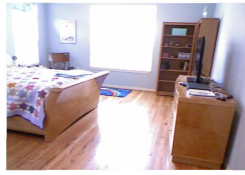
COCOQA 11372
What do two women hold with a picture on it?

Ground truth: cake
 IMG+BOW: cake (0.19)
 2-VIS+BLSTM: cake (0.19)
 BOW: umbrella (0.15)



DAQUAR 3018
What is on the right side?

Ground truth: table
 IMG+BOW: tv (0.28)
 2-VIS+LSTM: sofa (0.17)
 LSTM: cabinet (0.22)



DAQUAR 1426
What is on the right side table?

Ground truth: tv
 IMG+BOW: tv (0.25)
 2-VIS+LSTM: tv (0.29)
 LSTM: tv (0.14)

DAQUAR 1426a
What is on the left side of the room?

Ground truth: bed
 IMG+BOW: door (0.19)
 2-VIS+LSTM: door (0.25)
 LSTM: door (0.13)



COCOQA 15756
What does the man rid while wearing a black wet suit?

Ground truth: surfboard
 IMG+BOW: jacket (0.35)
 2-VIS+LSTM: surfboard (0.53)
 BOW: tie (0.30)



COCOQA 9715
What is displayed with the mattress off of it?

Ground truth: bed
 IMG+BOW: bench (0.36)
 2-VIS+LSTM: bed (0.18)
 BOW: airplane (0.08)



COCOQA 25124
What is sitting in a sink in the rest room?

Ground truth: cat
 IMG+BOW: toilet (0.77)
 2-VIS+LSTM: toilet (0.90)
 BOW: cat (0.83)

Figure 2: Sample questions and responses on Object questions



COCOQA 35
How many red velvet cup cakes with no frosting on a flowered plate?

Ground truth: three
 IMG+BOW: three (0.43)
 2-VIS+BLSTM: three (0.26)
 BOW: two (0.26)



COCOQA 21446
How many different types of carrying cases in various colors?

Ground truth: four
 IMG+BOW: four (0.69)
 2-VIS+BLSTM: four (0.26)
 BOW: three (0.25)



COCOQA 25600
How many double deckered busses parked near the green truck?

Ground truth: three
 IMG+BOW: two (0.89)
 2-VIS+BLSTM: two (0.44)
 BOW: two (0.64)



DAQUAR 555
How many chairs are there?

Ground truth: one
 IMG+BOW: six (0.20)
 2-VIS+BLSTM: one (0.20)
 LSTM: four (0.19)

Figure 3: Sample questions and responses on Number questions

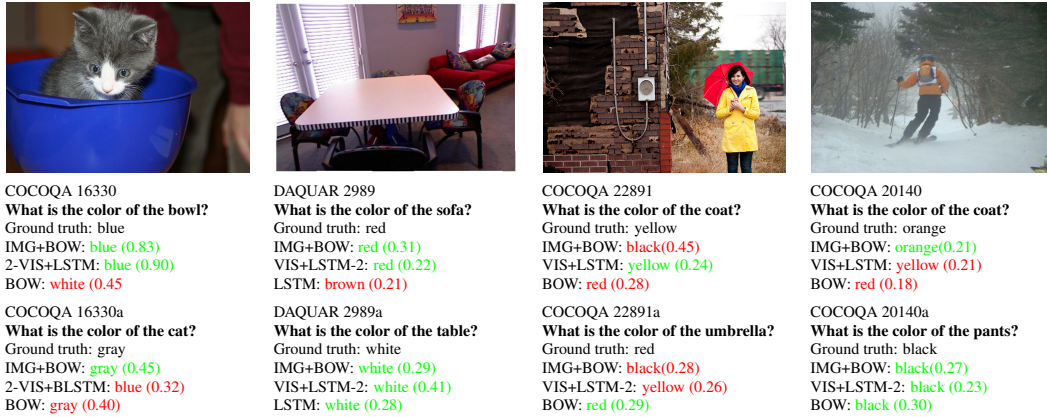


Figure 4: Sample questions and responses on Color questions

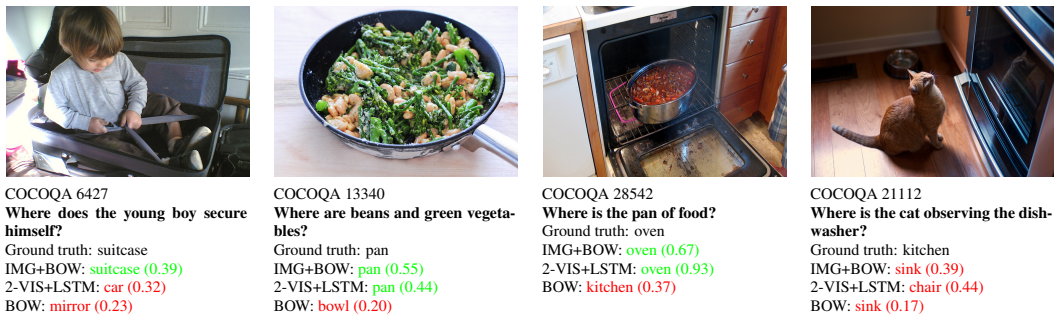


Figure 5: Sample questions and responses on Location questions