
Supplementary Material: Teaching Machines to Describe Images with Natural Language Feedback

David S. Hippocampus*
Department of Computer Science
Cranberry-Lemon University
Pittsburgh, PA 15213
hippo@cs.cranberry-lemon.edu

1 Feedback Crowd-Sourcing Interface

Please correct the most major mistake in the caption. If the caption is wrong in many places, we are hoping you would consider correcting more than one mistake (click "I want to do one more correction" once you finish with the first mistake). But you are allowed to submit with only one correction (click "go to evaluation"). See instructions for details.



(a cat) (sitting) (on a sidewalk) (next to a street .)

Correction Task:

Select type of mistake:

something should be replaced in the caption

Describe a mistake in natural language (write a full sentence(s) describing a mistake and suggest a correction):

There is a dog on a sidewalk, not a cat.

Select what is wrong:

wrong object

Select mistaken words:(please just input one wrong object/action/relation)

(select the mistaken word(s) by highlighting it with a mouse then click "select mistake words".

(a [Before]cat[After]) (sitting) (on a sidewalk) (next to a street .)

select mistake words

reset selection

Please correct the word(s) between your selected area.

(a [Before]dog[After]) (sitting) (on a sidewalk) (next to a street .)

Figure 1: Our web-based feedback collection interface.

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

2 Examples of Collected Feedback

Image	Before Correction	Feedback	After Correction
	(a man) (holding a hot dog) (in a restaurant .)	The man is wearing a red sweater and this should be mentioned.	(a man wearing a red sweater) (holding a hot dog) (in a restaurant .)
	(a living room) (with a bed) (and a television .)	There is no bed, only a couch.	(a living room) (with a couch) (and a television .)
	(a man) (sitting) (on a bench) (with a cat) (on the bench .)	there is a cup, but no cat on the bench.	(a man) (sitting) (on a bench) (with a cup) (on the bench .)
	(a bird) (is flying) (in the air) (with a frisbee .)	There is no frisbee in the picture	(a bird) (is flying) (in the air .)

Table 1: Examples of Collected Feedback

3 Qualitative Examples

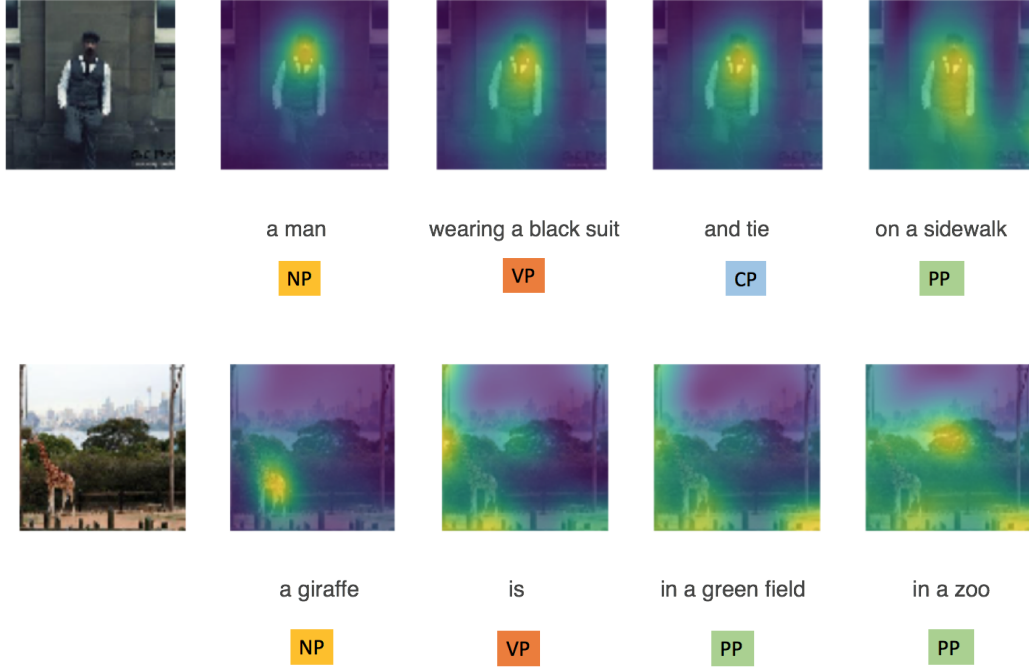


Figure 2: Examples of our phrase-based attention and phrase-label prediction.



Image	Caption
	<ul style="list-style-type: none"> • MLE: (a computer) (sitting) (on top of a desk) (on a monitor .) • RLB: (a laptop) (sitting) (on top of a desk) (next to a computer monitor .) • RLF: (a computer) (sitting) (on top of a desk) (with a monitor .)
	<ul style="list-style-type: none"> • MLE: (a suitcase) (is) (on a bed) (with a bag) (on top of it .) • RLB: (a suitcase) (is) (on a table) (with a suitcase .) • RLF: (a luggage bag) (sitting) (on a floor) (in a room .)

Table 2: Qualitative captioning results for our model and the baselines.



Image	Caption
	<ul style="list-style-type: none"> • MLE: (a red bus) (driving down a street) (with a person) (waiting) (on the street .) • RLB:(a red bus) (driving down a street) (with people) (driving) (on the side .) • RLF:(a red bus) (is driving down the street .)
	<ul style="list-style-type: none"> • MLE:(a street) (with a traffic light) (and a bus) (in the background .) • RLB:(a person) (walking) (on a city street) (with a yellow sign .) • RLF: (a street) (with a car) (is driving down a street .)

Table 3: caption results


Image	Caption
	<ul style="list-style-type: none"> • MLE: (a man) (is jumping) (into the air to) (catch a frisbee .) • RLB: (a man) (is throwing a frisbee) (in a park .) • RLF: (a man) (is holding a frisbee) (in his mouth .)

Table 4: An example of a failed caption, where however the baselines produce a more reasonable caption.

4 References