

---

# Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

## (Supplementary Material)

---

**Kelvin Xu**

Université de Montréal

KELVIN.XU@UMONTREAL.CA

**Jimmy Lei Ba**

University of Toronto

JIMMY@PSI.UTORONTO.CA

**Ryan Kiros**

University of Toronto

RKIROS@CS.TORONTO.EDU

**Kyunghyun Cho**

Université de Montréal

KYUNGHYUN.CHO@UMONTREAL.CA

**Aaron Courville**

Université de Montréal

AARON.COURVILLE@UMONTREAL.CA

**Ruslan Salakhutdinov**

University of Toronto

RSALAKHU@CS.TORONTO.EDU

**Richard S. Zemel**

University of Toronto

ZEMEL@CS.TORONTO.EDU

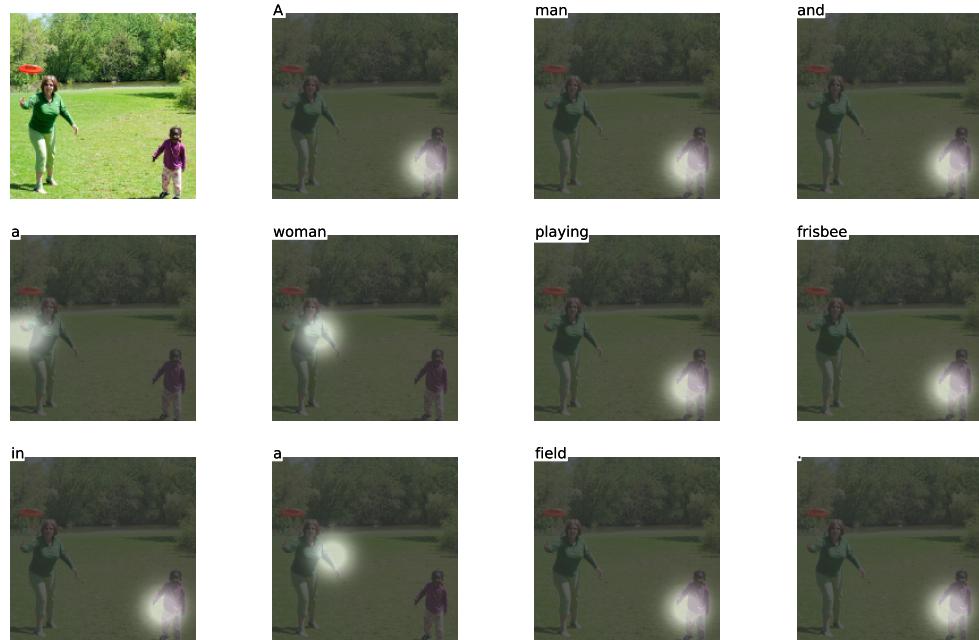
**Yoshua Bengio**

Université de Montréal

YOSHUA.BENGIO@UMONTREAL.CA

## 1. Additional Visualizations

Visualizations from our “hard” (a) and “soft” (b) attention model. *White* indicates the regions where the model roughly attends to.

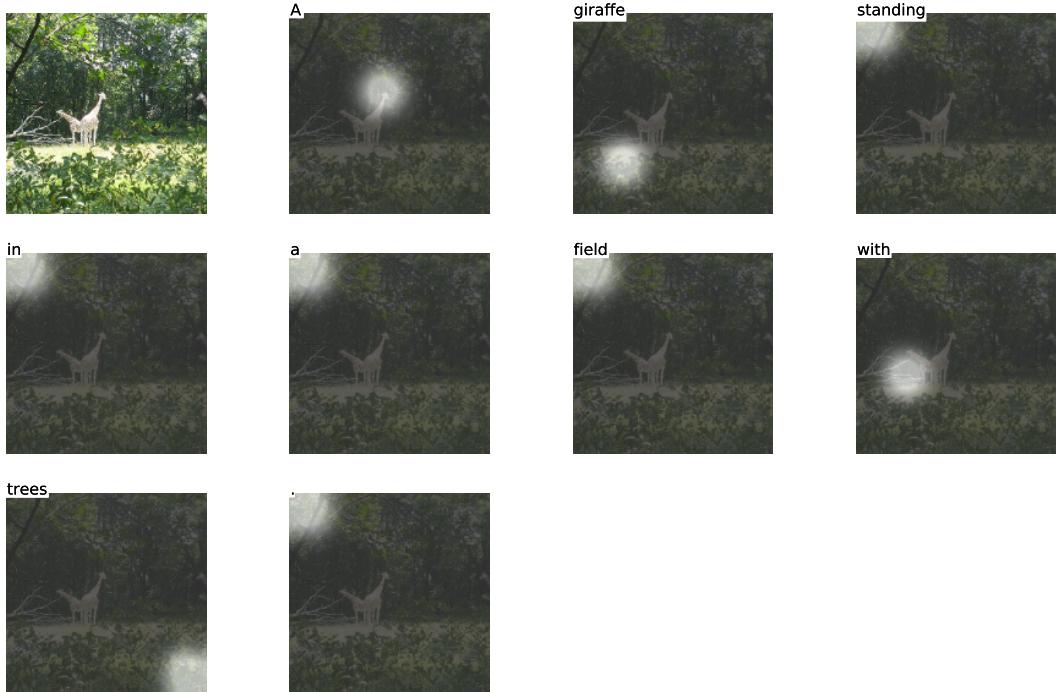


(a) A man and a woman playing frisbee in a field.

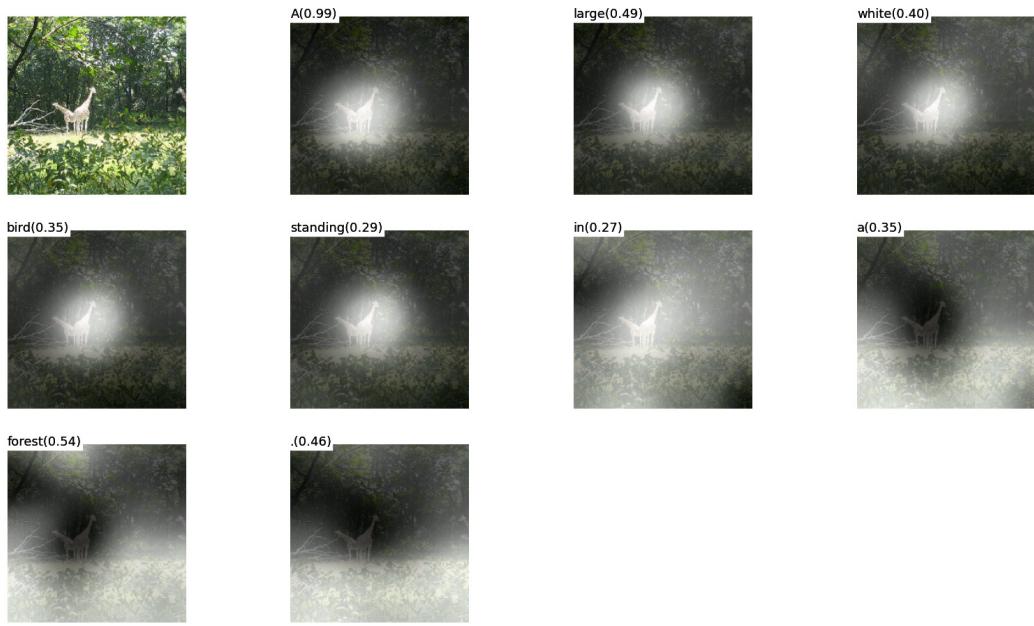


(b) A woman is throwing a frisbee in a park.

Figure 1.

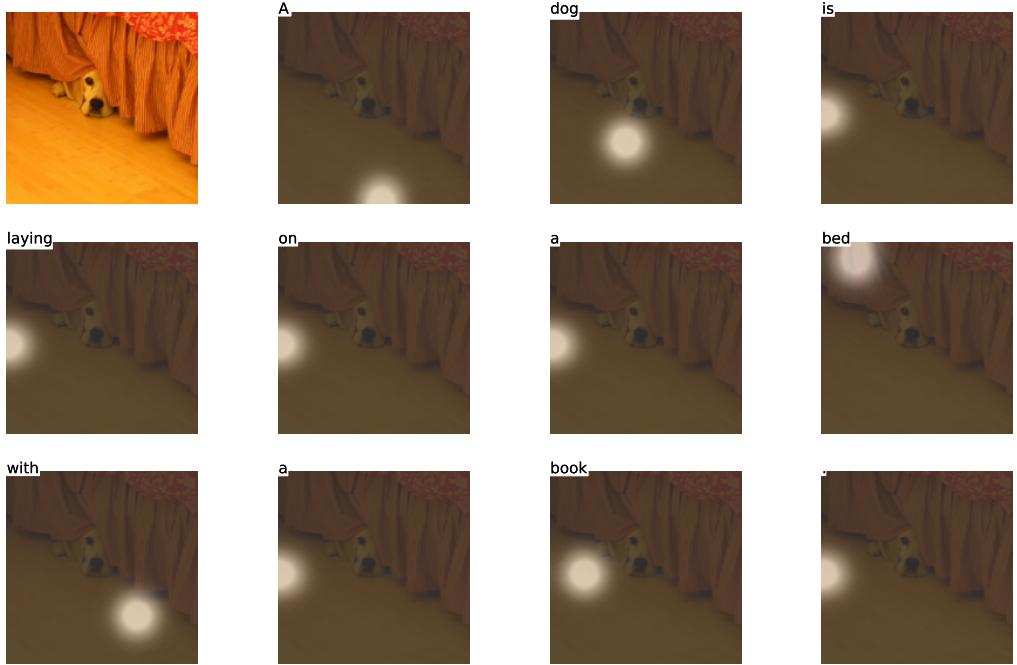


(a) A giraffe standing in the field with trees.

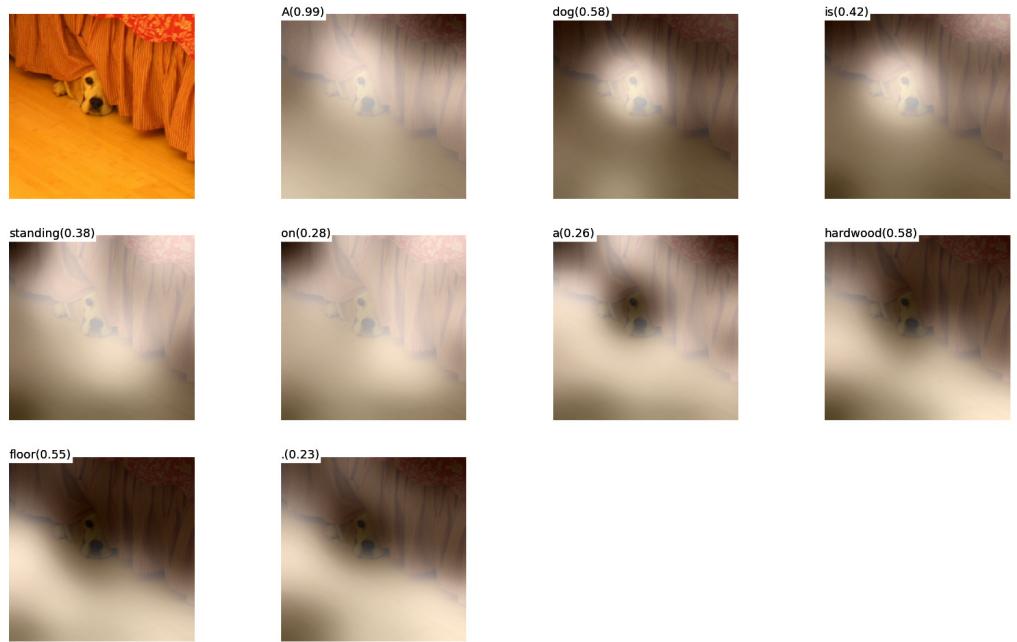


(b) A large white bird standing in a forest.

*Figure 2.*

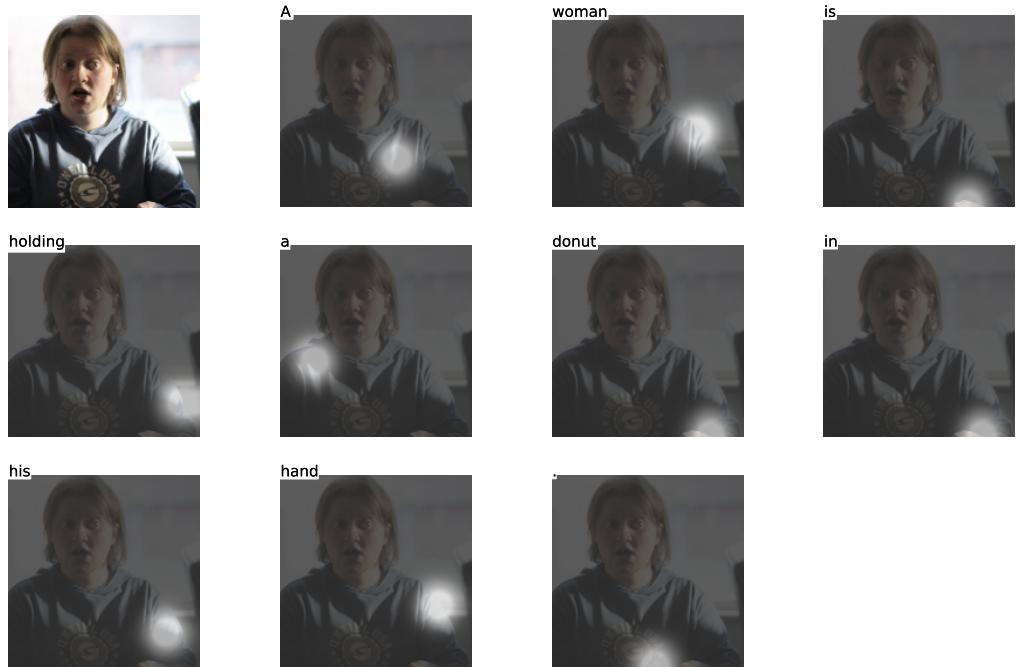


(a) A dog is laying on a bed with a book.



(b) A dog is standing on a hardwood floor.

*Figure 3.*

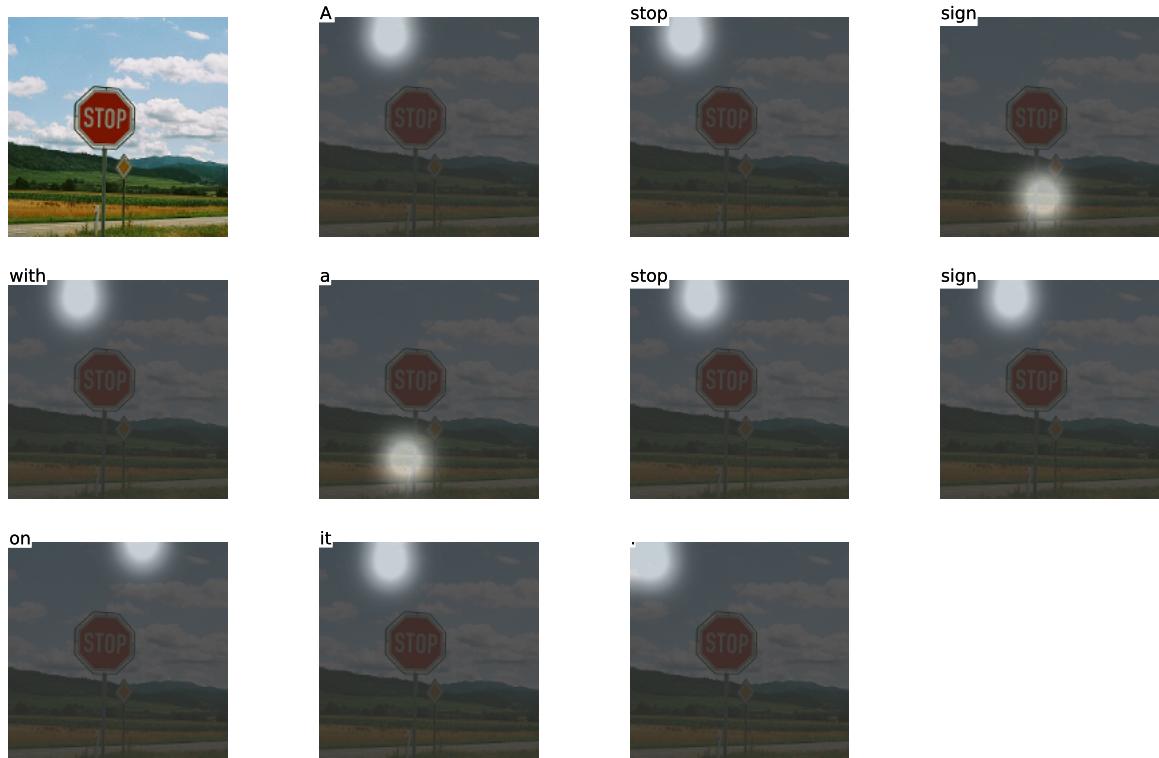


(a) A woman is holding a donut in his hand.

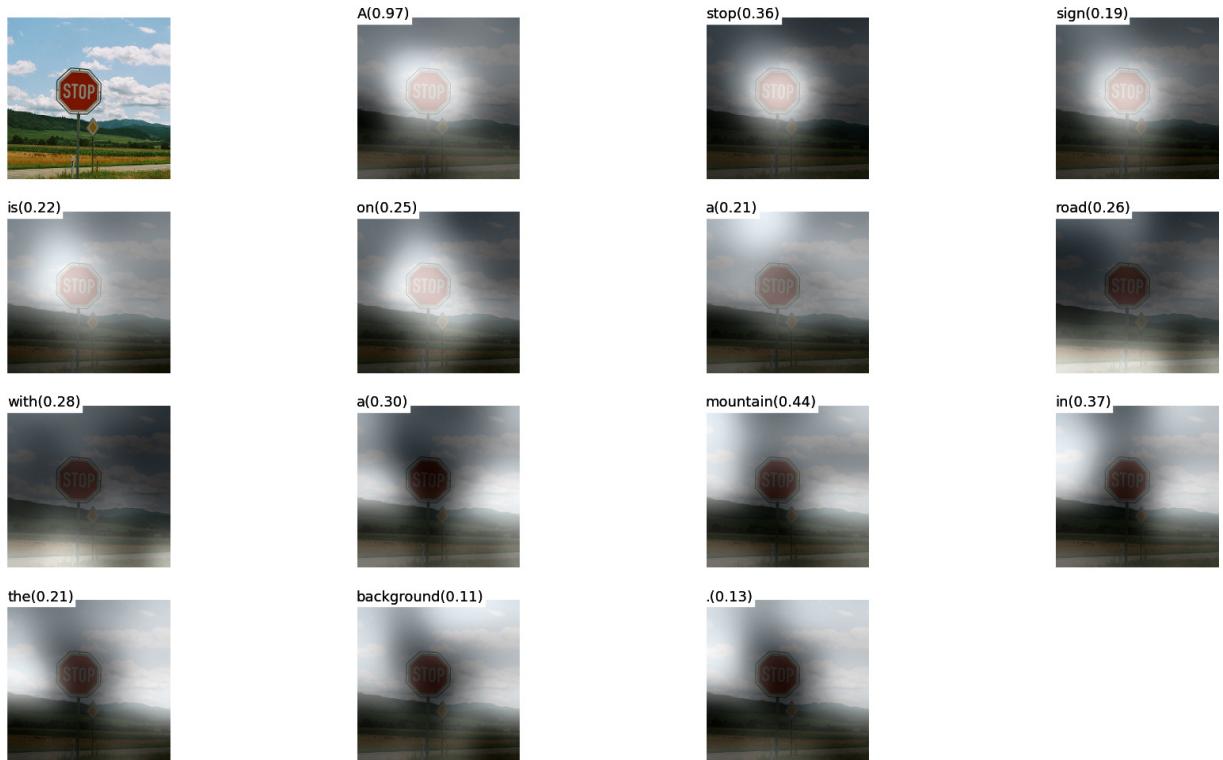


(b) A woman holding a clock in her hand.

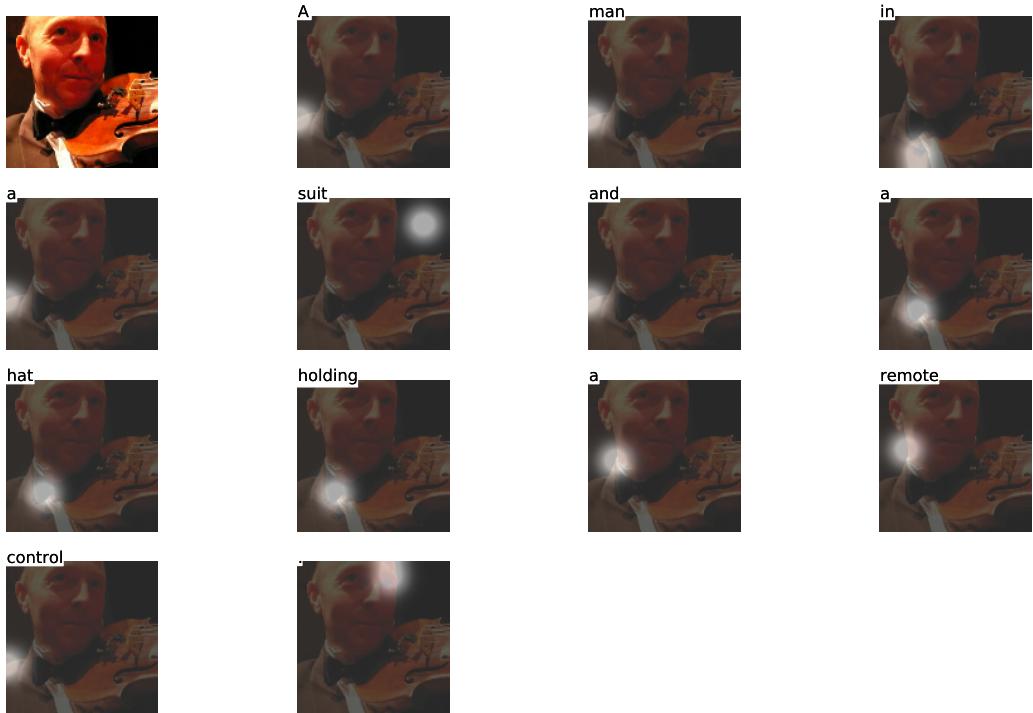
Figure 4.



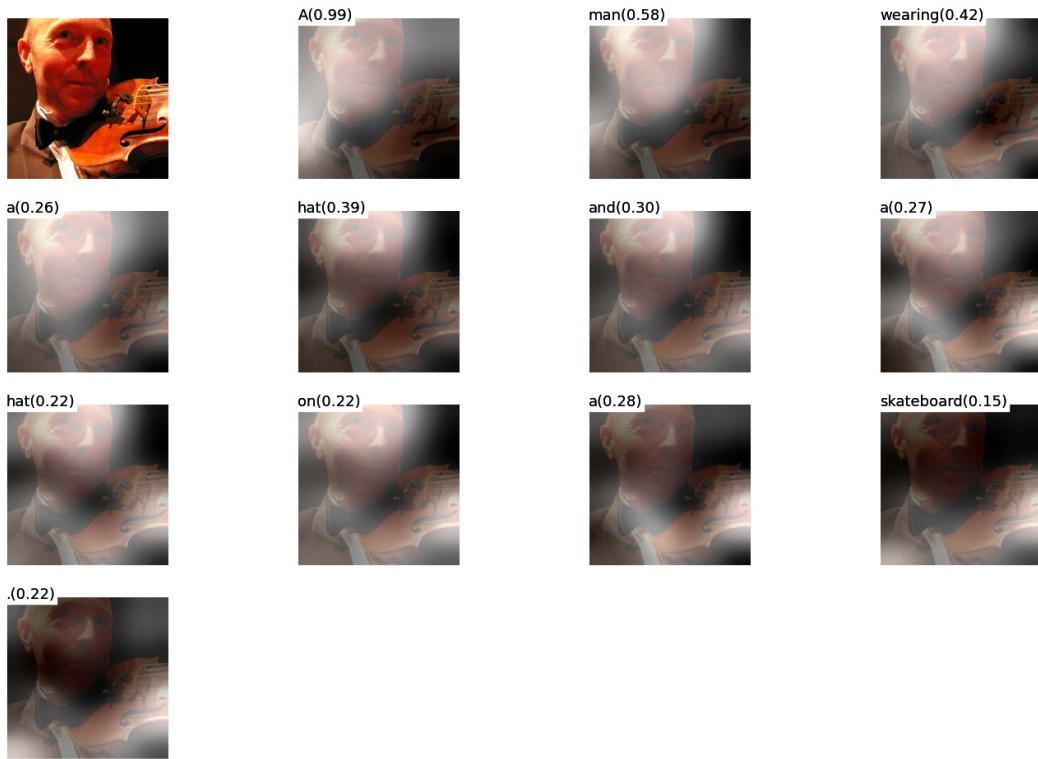
(a) A stop sign with a stop sign on it.



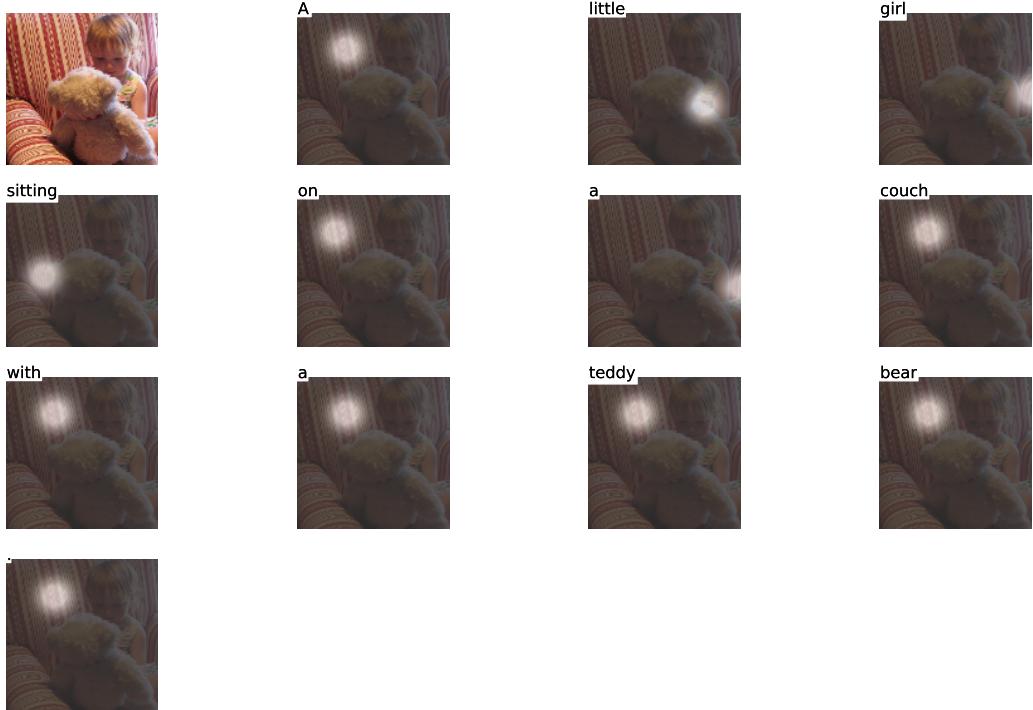
(b) A stop sign is on a road with a mountain in the background.



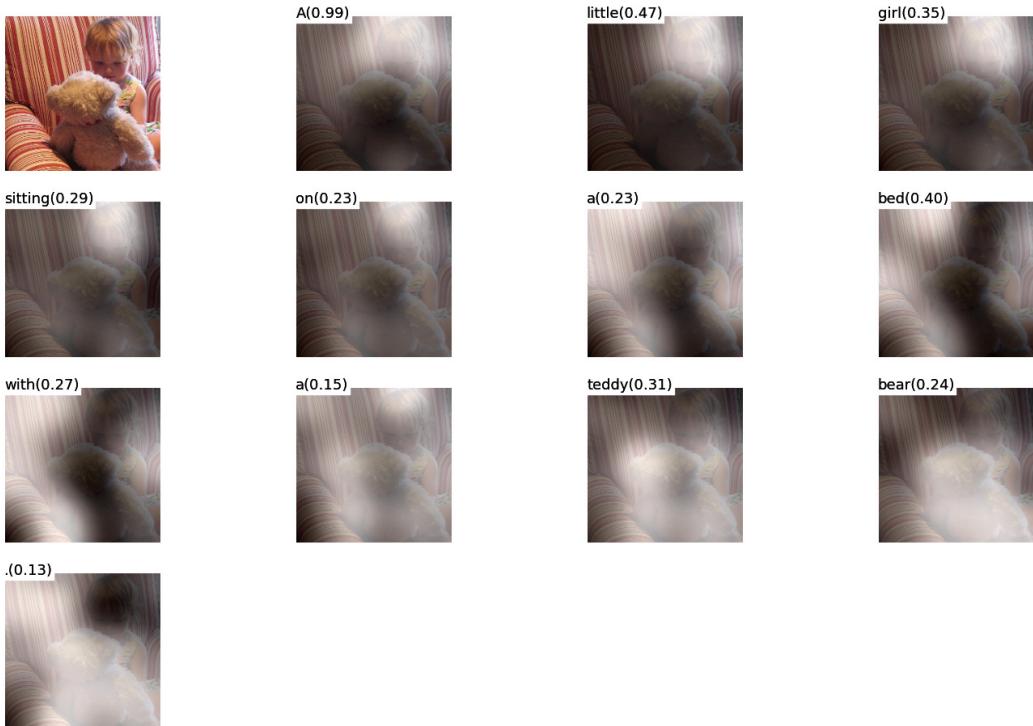
(a) A man in a suit and a hat holding a remote control.



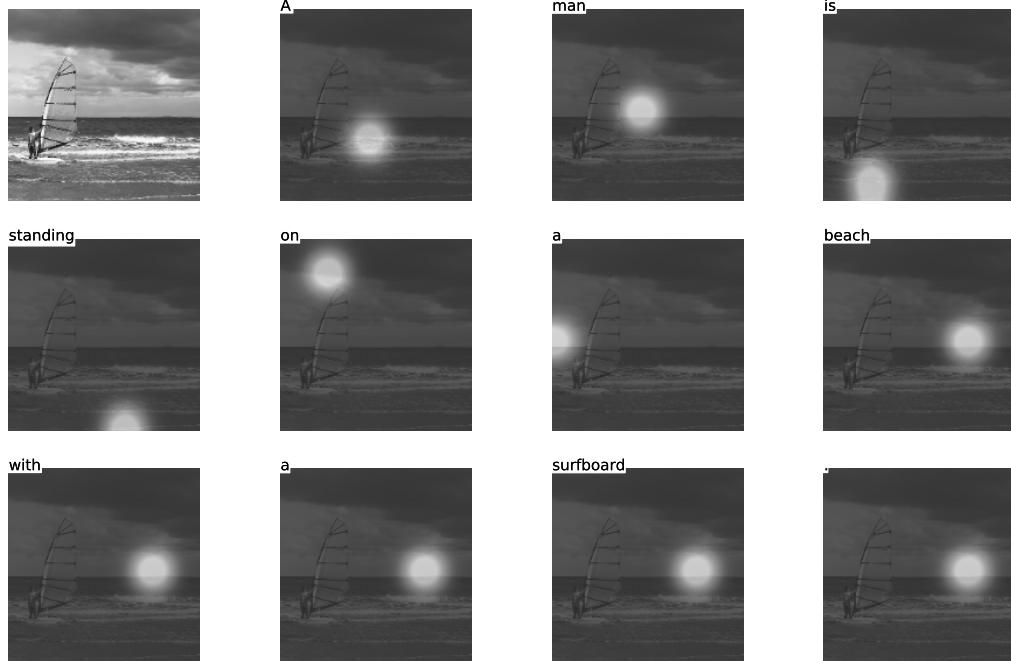
(b) A man wearing a hat and a hat on a skateboard.



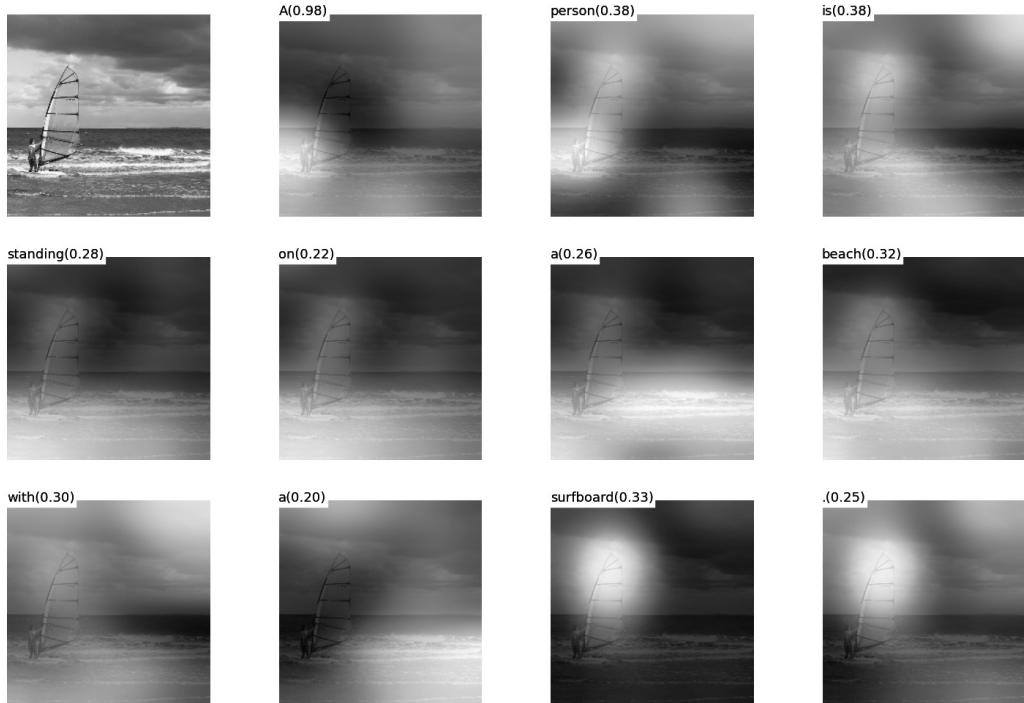
(a) A little girl sitting on a couch with a teddy bear.



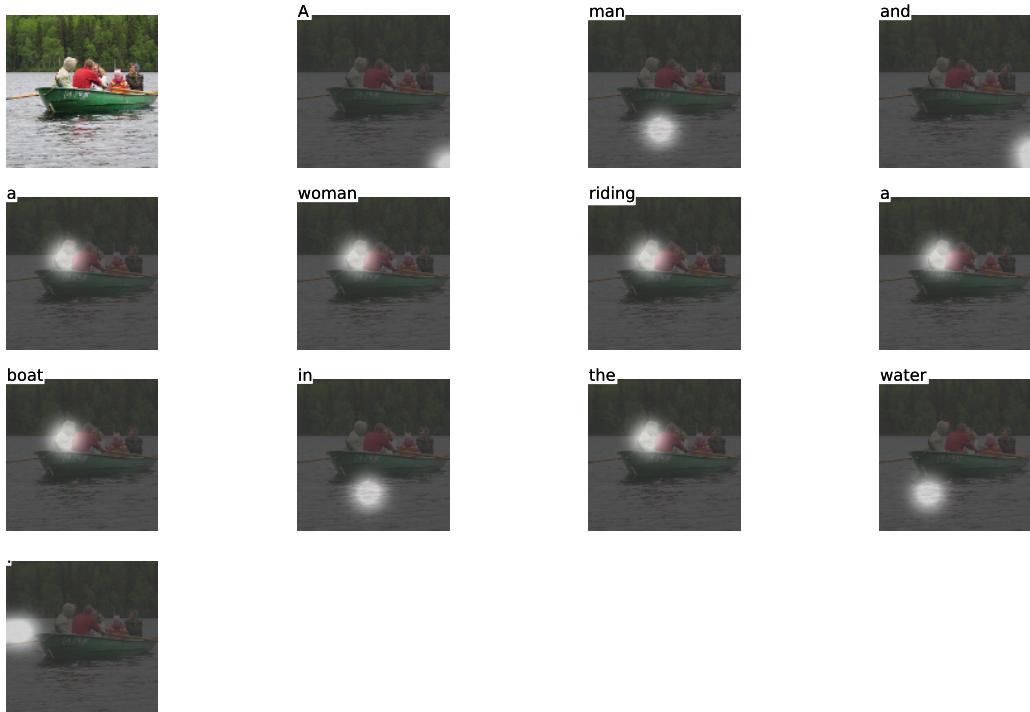
(b) A little girl sitting on a bed with a teddy bear.



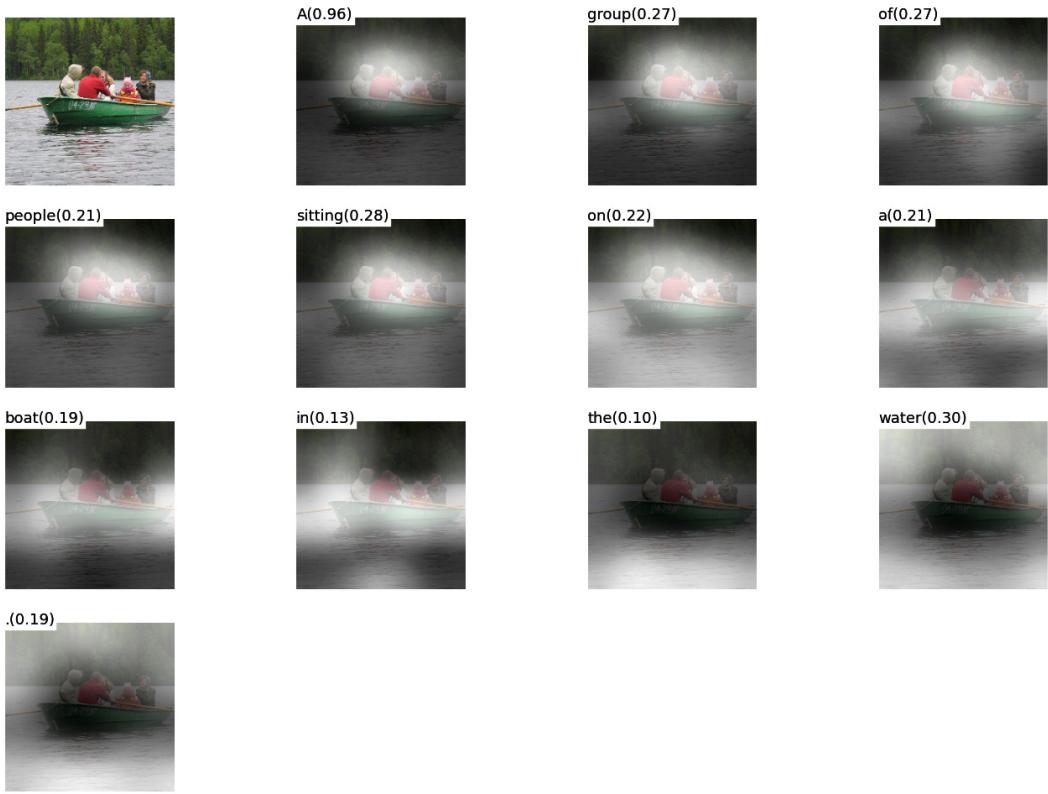
(a) A man is standing on a beach with a surfboard.



(b) A person is standing on a beach with a surfboard.

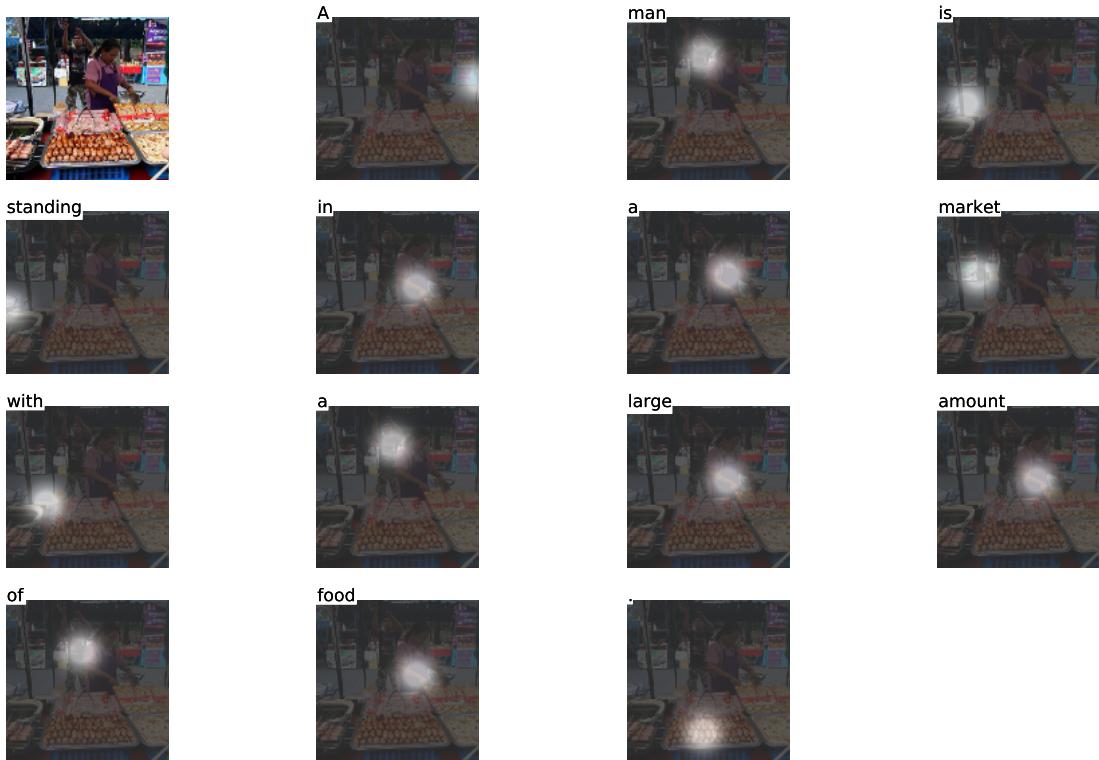


(a) A man and a woman riding a boat in the water.

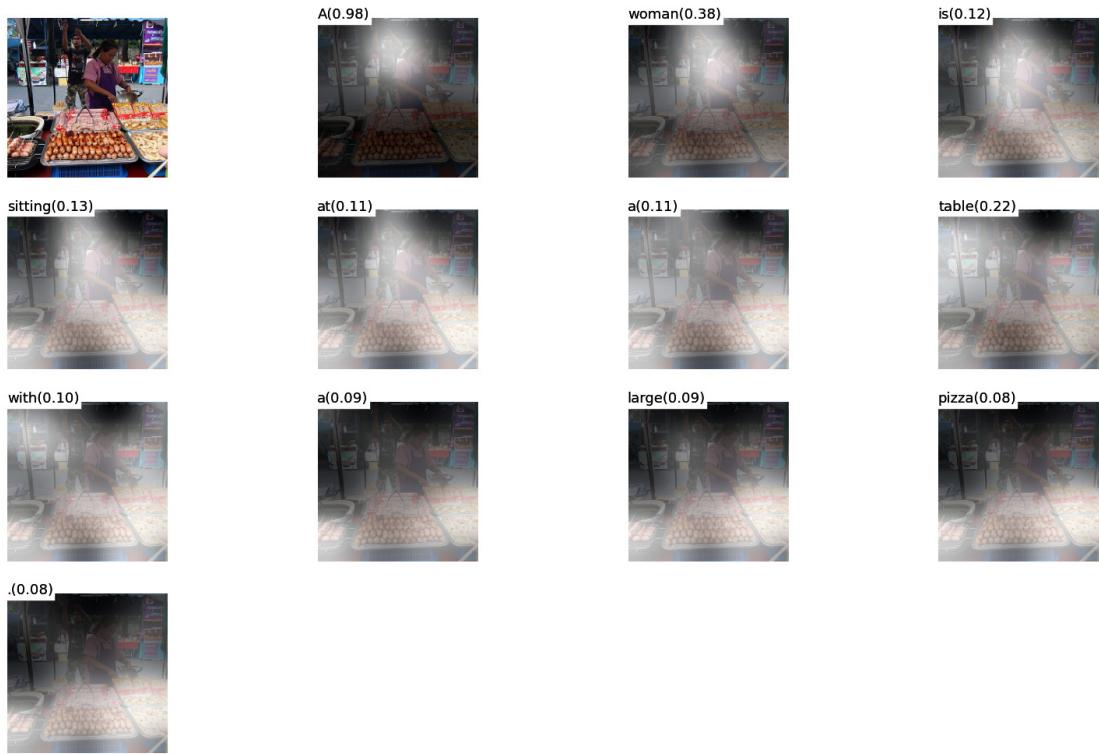


(b) A group of people sitting on a boat in the water.

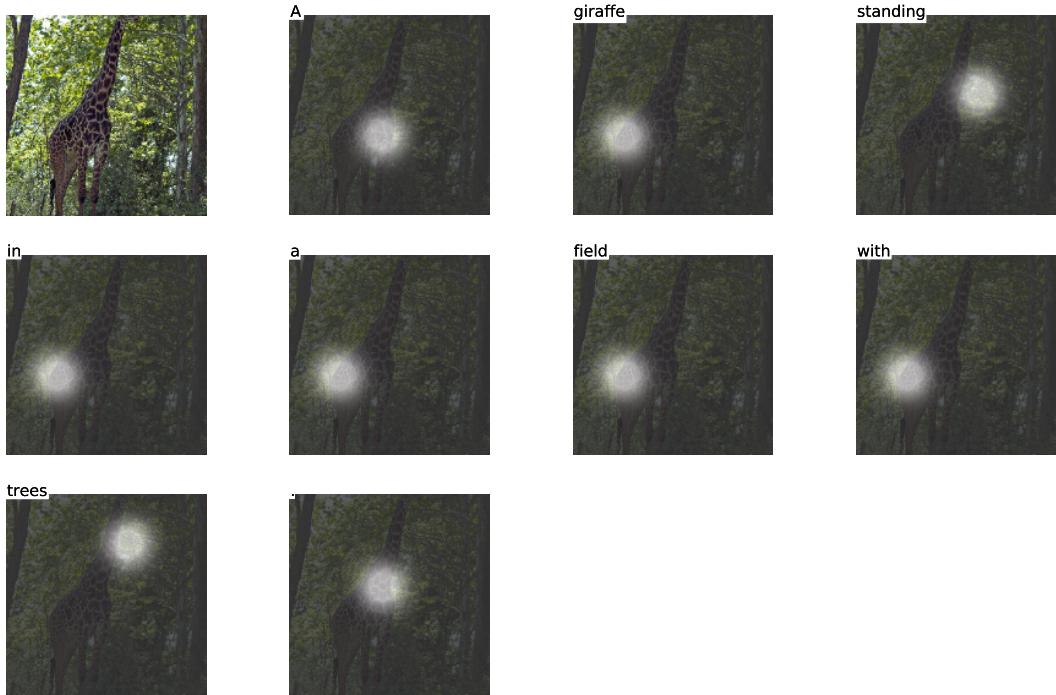
Figure 7.



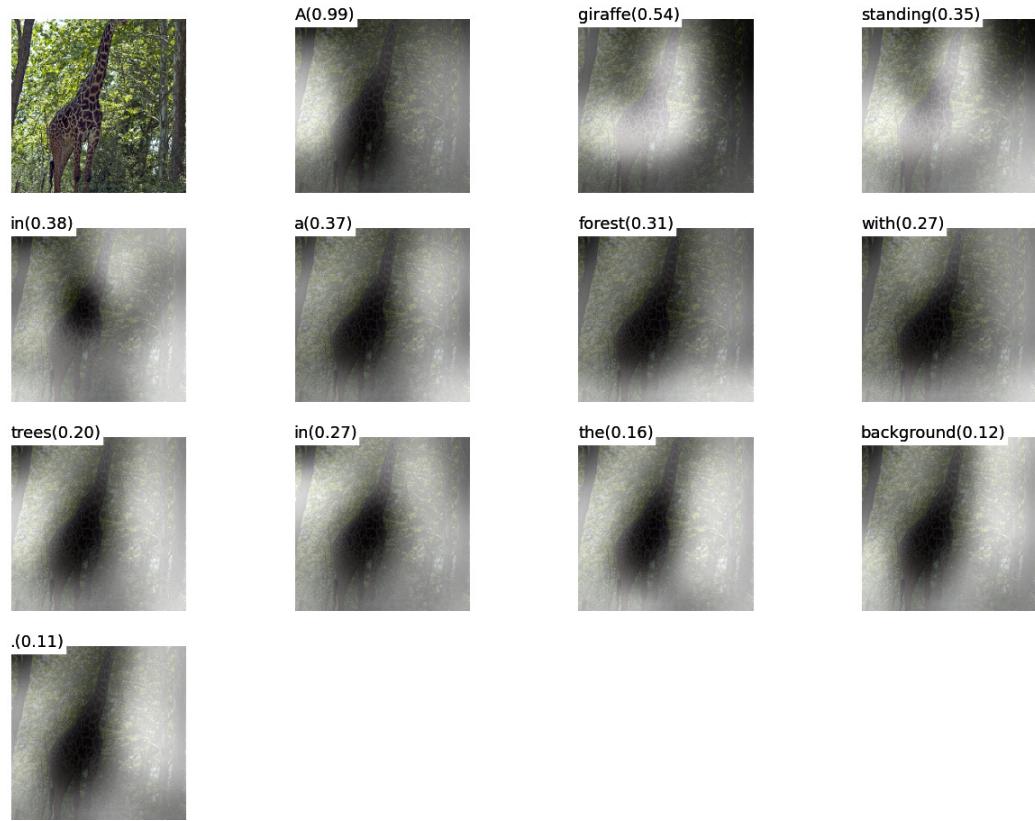
(a) A man is standing in a market with a large amount of food.



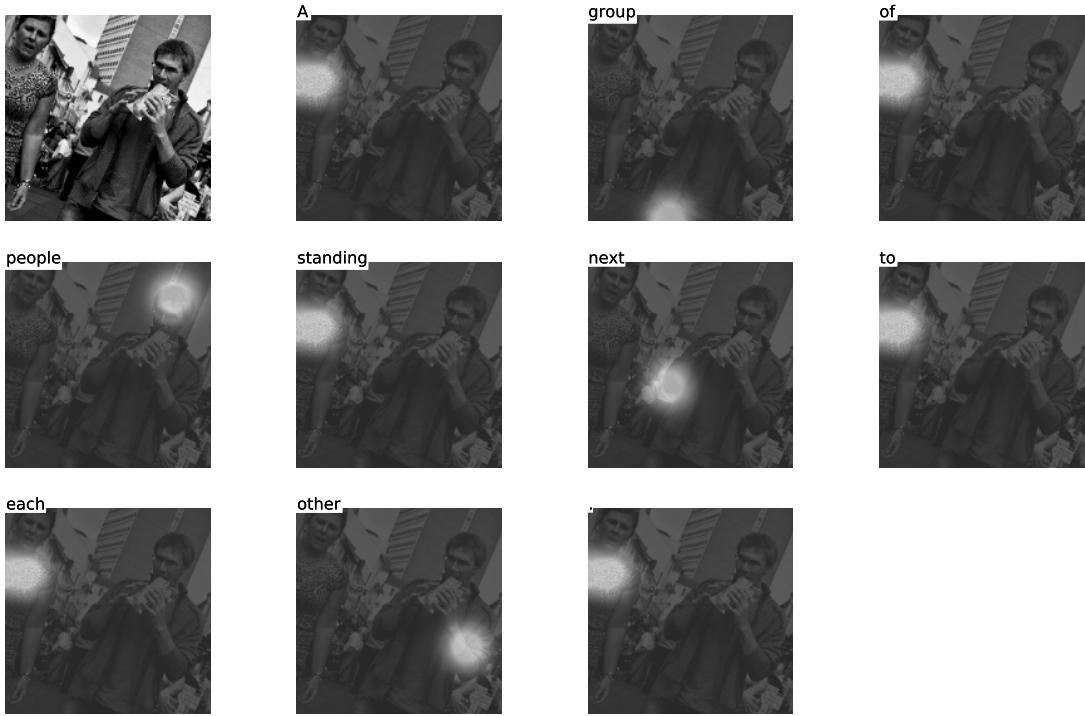
(b) A woman is sitting at a table with a large pizza.



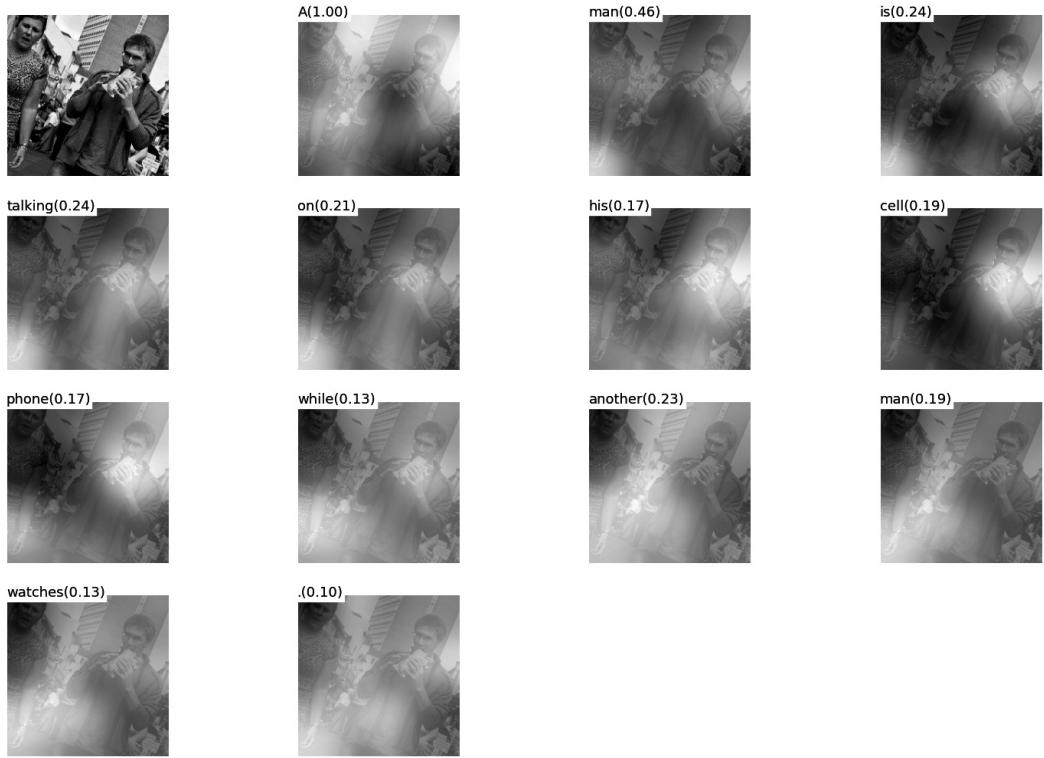
(a) A giraffe standing in a field with trees.



(b) A giraffe standing in a forest with trees in the background.



(a) A group of people standing next to each other.



(b) A man is talking on his cell phone while another man watches.