

Visual Madlibs: Fill in the blank Description Generation and Question Answering

Supplementary File

Licheng Yu, Eunbyung Park, Alexander C. Berg, Tamara L. Berg
Department of Computer Science, University of North Carolina, Chapel Hill
{licheng, eunbyung, aberg, tlberg}@cs.unc.edu

1. Results of filtered easy and hard tasks

We show the full tables of accuracies for the filtered easy and hard multiple-choice tasks in Table 1.

2. Quantitative analysis of Madlibs responses

In Section 5.1 of our paper, we analyzed the the phrasal structures of our collected Visual Madlibs descriptions for several of our fill-in-the-blank questions. Here, we show the relative frequencies for the top-5 most frequent templates used for all 12 Madlibs questions. In Fig. 1, it is observed that most of the distributions are concentrated on just a few choices, except for the future and past descriptions. One reason is that this question is more open ended, so Turkers are likely to write more lengthy descriptions, i.e., “One or two seconds before this picture was taken, ___”. In this setup, annotators have great freedom in expressing their ideas.

We also show an analysis of answer consistency in Fig. 2. Here we compute a histogram of answer similarities for each question, where similarity is measured as the cosine similarity between the mean Word2Vec representations of the 3 collected answers. Most of the similarity histograms have a normal-like distribution with an extra peak around 1, which implies we get some very similar answers for a portion of questions. The exception is the answers for the questions about the past and future, where there is no peak at one for distribution of similarities. This indicates that there are fewer images for past/future predictions where people generate the exact same description, but for some images people do display more consistency than for other images. Also note, the mean for image’s emotion is smaller compared with the others, perhaps indicating that this question is relatively more subjective.

Finally, we analyze the word distribution for each question type. In Fig. 3, we show the top 20 most frequently used words for all types of Madlibs questions. It is interesting to find that color words are more often used to

describe object’s attribute, while the entry-level category words (‘woman’, ‘boy’, ‘girl’, ‘child’) and the clothes are usually used to describe person’s attribute. More than 20% of the images express positive emotions, i.e., ‘happy’ and ‘excited’. Several words related to person are often used to indicate the relative position of object, e.g, ‘hand’, ‘person’, ‘man’, ‘woman’, etc. Perhaps it is due to the human-centric property of the Visual Madlibs images.

3. Additional examples of results

We also show additional examples of the two Visual Madlibs tasks: multiple-choice question answering and focused sentence generation. We first show some correct answers to multiple choice questions in Fig. 4, as well as some wrong answers in Fig. 5. All examples are from the hard version of our multiple-choice question answering task and answers are selected by the nCCA joint-embedding method. This task provides a more straightforward way to measure the quality of the learned joint embedding space in an application scenario.

Then, in Fig. 6, we show some focused sentence generation examples, generated by nCCA and CNN+LSTM. As observed, the nCCA can generate richer but sometimes unrelated sentences, while CNN+LSTM is able to generate relatively shorter but accurate sentences, which helps to achieve higher BLEU-1 and BLEU-2 scores.

Filtered Questions from Easy Task										
	#Q	n-gram	CCA	nCCA	nCCA (place)	nCCA (bbox)	nCCA (all)	CNN+LSTM (madlibs)	CNN+LSTM(r) (madlibs)	Human
1. scene	5997	24.6%	77.4%	88.8%	87.4%	—	89.7%	76.1%	79.4%	96.4%
2. emotion	2663	27.5%	48.3%	58.8%	59.7%	—	51.0%	39.4%	49.0%	75.5%
3. past	4703	26.0%	62.8%	78.9%	73.9%	—	81.7%	50.5%	47.1%	96.8%
4. future	4495	28.7%	62.2%	79.4%	73.3%	—	81.5%	51.2%	51.4%	97.1%
5. interesting	4940	26.4%	67.6%	77.5%	72.9%	—	79.8%	56.1%	50.5%	96.8%
6. obj attr	6681	32.2%	45.1%	48.9%	45.8%	56.6%	52.4%	48.3%	60.8%	93.3%
7. obj aff	7043	31.0%	60.8%	74.3%	70.8%	73.5%	77.9%	—	90.8%	95.8%
8. obj pos	6906	27.0%	54.1%	67.3%	65.6%	60.3%	71.0%	54.9%	71.5%	94.9%
9. per attr	5753	27.3%	42.1%	50.5%	46.6%	56.6%	46.2%	37.2%	49.2%	92.2%
10. per act	6384	27.3%	70.6%	81.2%	77.4%	76.3%	83.3%	65.1%	69.5%	97.9%
11. per loc	6193	24.7%	72.0%	85.1%	85.2%	76.1%	85.3%	62.1%	73.6%	96.8%
12. pair rel	7206	29.5%	55.4%	64.4%	62.8%	65.6%	68.6%	—	74.0%	95.1%

Filtered Questions from Hard Task										
	#Q	n-gram	CCA	nCCA	nCCA (place)	nCCA (bbox)	nCCA (all)	CNN+LSTM (madlibs)	CNN+LSTM(r) (madlibs)	Human
1. scene	4940	20.9%	70.4%	77.6%	77.8%	—	76.3%	69.4%	69.7%	89.5%
2. emotion	2052	27.7%	43.1%	49.0%	49.5%	—	43.8%	38.5%	43.0%	72.2%
3. past	3976	24.3%	51.0%	57.4%	53.8%	—	59.4%	42.8%	41.3%	86.6%
4. future	3820	25.7%	51.4%	59.2%	54.2%	—	58.3%	42.1%	41.7%	87.6%
5. interesting	4159	26.9%	56.1%	59.5%	55.1%	—	61.3%	47.8%	40.3%	89.5%
6. obj attr	5436	30.3%	45.3%	47.2%	44.7%	54.6%	42.8%	45.1%	46.3%	86.2%
7. obj aff	4581	28.0%	61.2%	71.0%	67.6%	70.5%	57.6%	—	79.0%	73.7%
8. obj pos	5721	29.1%	53.0%	60.2%	57.7%	54.6%	57.7%	48.8%	54.3%	84.5%
9. per attr	4893	25.7%	36.5%	42.4%	38.8%	52.1%	34.4%	36.1%	46.4%	88.2%
10. per act	5813	27.7%	62.0%	68.3%	65.3%	67.9%	69.6%	59.1%	55.3%	92.7%
11. per loc	5096	23.6%	63.1%	69.9%	71.7%	62.6%	70.0%	52.9%	60.6%	88.2%
12. pair rel	5981	28.5%	52.3%	57.6%	55.4%	60.0%	56.5%	—	57.4%	88.5%

Table 1: Accuracies computed for different approaches on the filtered multiple-choice questions of easy and hard task. CCA, nCCA, and CNN+LSTM are trained on the whole image representation for each type of question. nCCA(place) uses Places-CNN feature. nCCA(box) is trained and evaluated on ground-truth bounding-boxes from COCO segmentations. nCCA(all) trains a single embedding using all question types. CNN+LSTM(r) ranks the perplexity of {prompt+choice}.

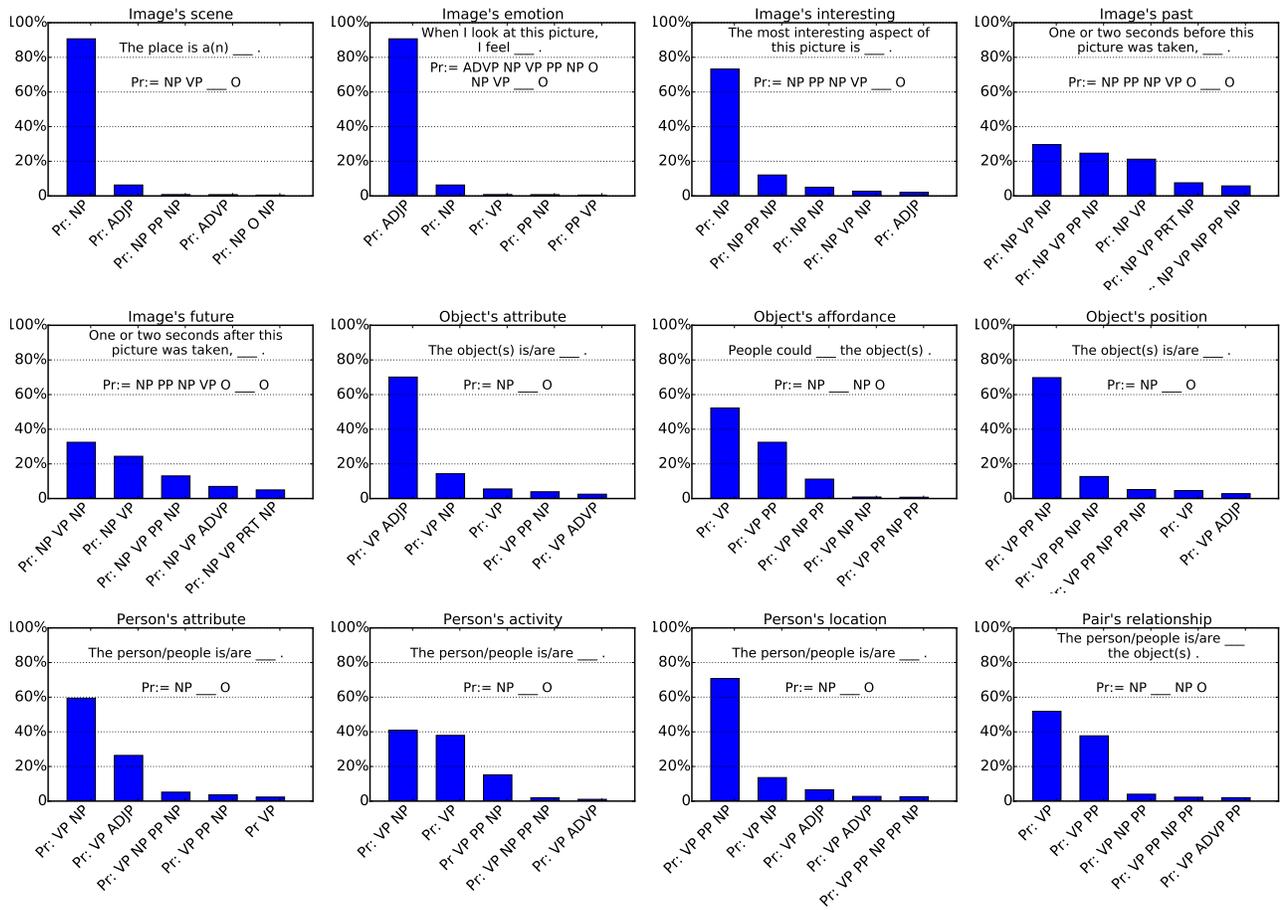


Figure 1: Top-5 most frequent phrase templates for 12 types of Madlibs questions.

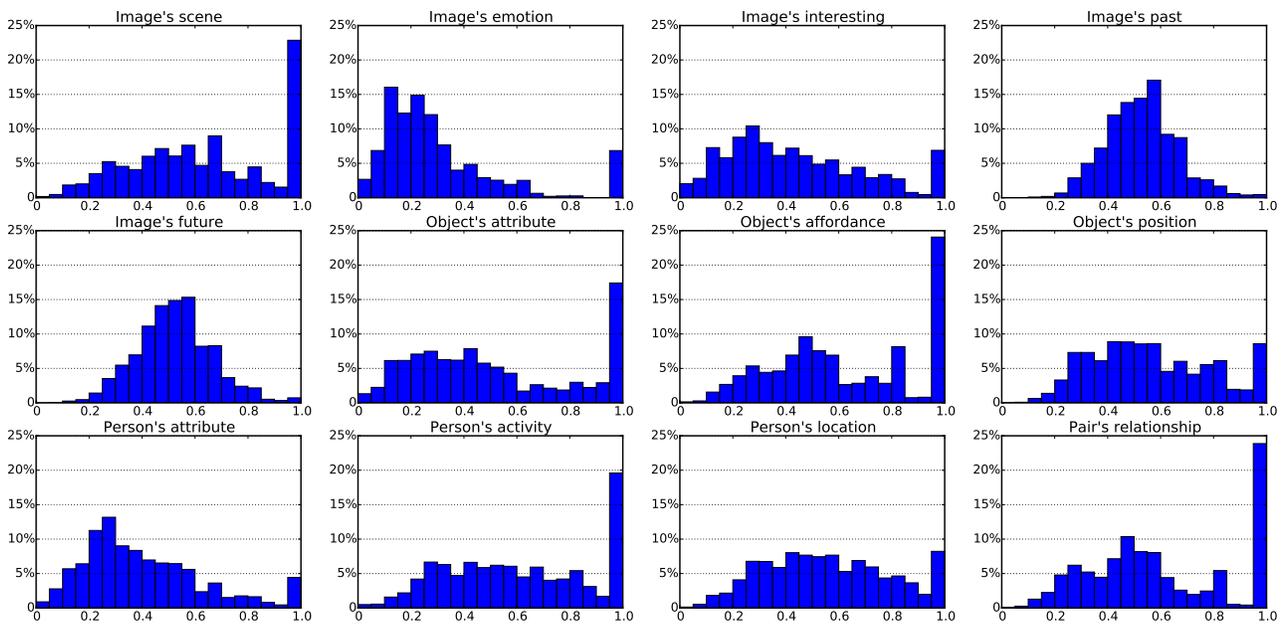


Figure 2: Histograms of similarity of answers for 12 types of Madlibs questions.

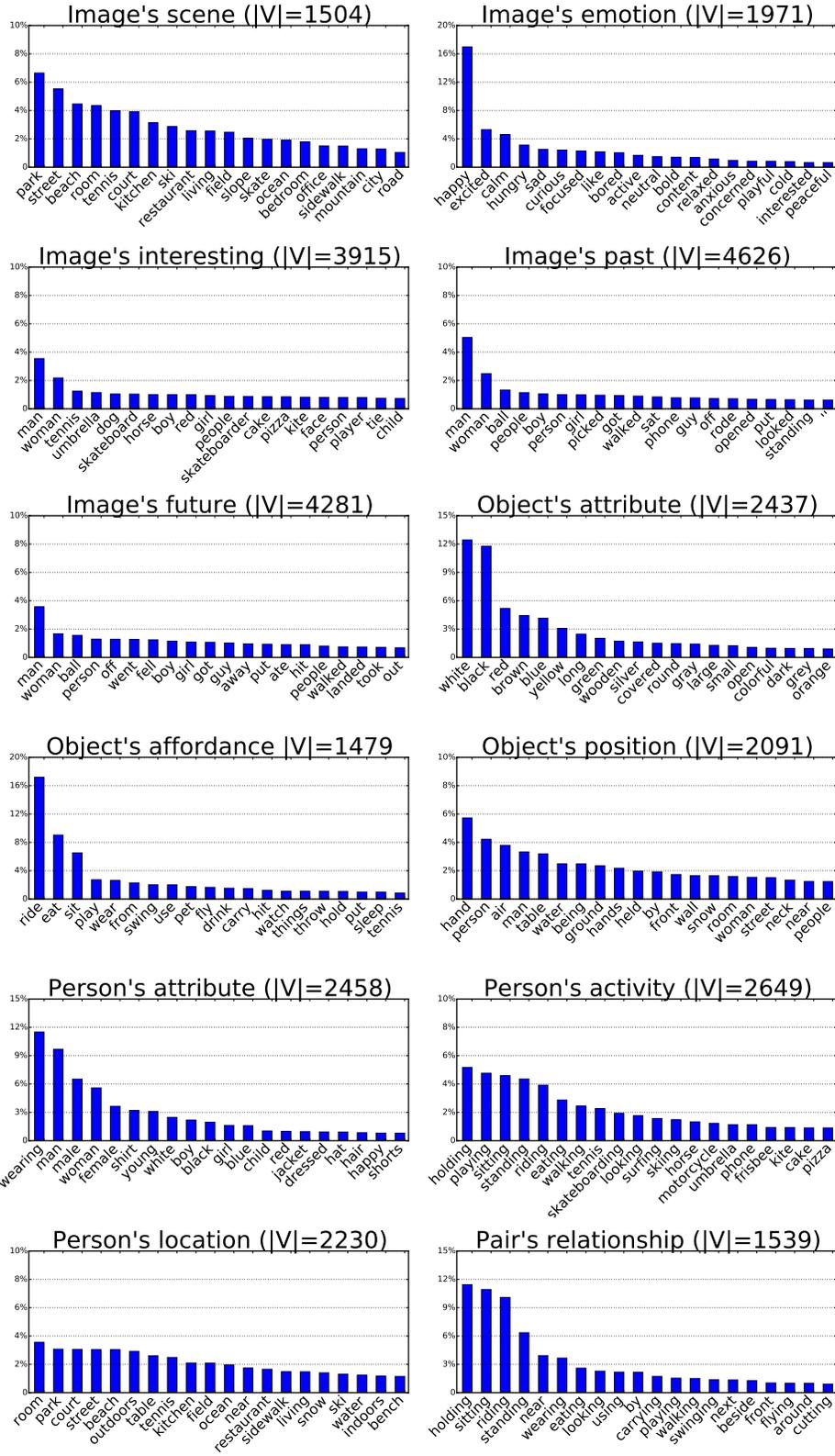


Figure 3: Top 20 words used in the answers for all 12 types of questions. |V| denotes the vocabulary size. Note y-axis range differs across question types, depending on the highest bin of each.

1. image's scene

 <p>The place is a(n) ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> mechanics shop <input type="checkbox"/> office <input type="checkbox"/> tennis match <input type="checkbox"/> living room 	 <p>The place is a(n) ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> track <input type="checkbox"/> lawn <input checked="" type="checkbox"/> park <input type="checkbox"/> tennis court 	 <p>The place is a(n) ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> office <input type="checkbox"/> park <input checked="" type="checkbox"/> tennis court <input type="checkbox"/> beach 	 <p>The place is a(n) ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> beach <input type="checkbox"/> ocean <input type="checkbox"/> riverfront <input type="checkbox"/> dock
---	---	--	--

2. image's emotion

 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> calm <input type="checkbox"/> happy <input type="checkbox"/> gross <input type="checkbox"/> stressed 	 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> relaxed <input type="checkbox"/> affection <input type="checkbox"/> happy <input type="checkbox"/> rejected 	 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> thirsty <input type="checkbox"/> mad <input checked="" type="checkbox"/> hungry <input type="checkbox"/> playful 	 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> bored <input type="checkbox"/> happy <input type="checkbox"/> peace <input checked="" type="checkbox"/> hot
--	---	--	--

3. image's interesting

 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the sheep with a pink head <input type="checkbox"/> dog <input type="checkbox"/> the facial expressions <input type="checkbox"/> the sheep's belly 	 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> horses racing <input type="checkbox"/> the rider <input type="checkbox"/> the huge watermark <input type="checkbox"/> the vinageness of this picture 	 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the boat in the background <input checked="" type="checkbox"/> the blue water is gorgeous <input type="checkbox"/> the man is about to fall over <input type="checkbox"/> the man 	 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the donuts <input type="checkbox"/> the girl's posture <input type="checkbox"/> the texture of the donut and the skin <input type="checkbox"/> the eye that can see through the the donuts
--	--	---	--

4. image's past

 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the man was surfing <input type="checkbox"/> the surfer gave the thumbs up <input type="checkbox"/> the woman put on her sunglasses <input type="checkbox"/> several people sat down 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> they cut the cake <input type="checkbox"/> the person asked another person to take a picture. <input checked="" type="checkbox"/> a man embraces a woman <input type="checkbox"/> tennis court 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the person was holding the frisbee <input type="checkbox"/> he jumped up <input type="checkbox"/> the family piled out of a truck <input type="checkbox"/> a wave came ashore 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the ball was hit to her side of the court <input type="checkbox"/> he picked up his racket <input type="checkbox"/> she received a serve <input type="checkbox"/> a man sat
--	---	---	---

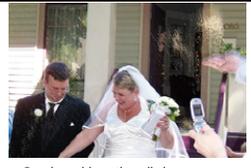
5. image's future

 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the guy was facing the camera <input checked="" type="checkbox"/> the woman takes a bit out of the banana <input type="checkbox"/> the little boy knocked over the pile <input type="checkbox"/> the man looked at the camera 	 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> someone make a toast <input type="checkbox"/> the man felt queasy <input checked="" type="checkbox"/> she ate the pizza <input type="checkbox"/> the girl stopped smiling 	 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> a girl cuts a piece of cake <input type="checkbox"/> the man will pick up a cup <input type="checkbox"/> the guy will get his own phone <input type="checkbox"/> the man talks babytalk to the baby 	 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the boy continued to smile <input type="checkbox"/> the boys will eat carrots <input type="checkbox"/> she will jump on the bed <input checked="" type="checkbox"/> a woman closed a laptop
---	---	---	---

6. object's attribute

 <p>The tennis racket is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> black <input type="checkbox"/> red <input type="checkbox"/> metal <input type="checkbox"/> plastic 	 <p>The skis is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> blue <input type="checkbox"/> colorful <input checked="" type="checkbox"/> long <input type="checkbox"/> short 	 <p>The surfboard is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> long and slim <input type="checkbox"/> red and black <input checked="" type="checkbox"/> white and blue <input type="checkbox"/> yellow and red 	 <p>The laptop is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> black <input type="checkbox"/> on <input type="checkbox"/> gray <input type="checkbox"/> silver
---	---	--	---

7. object's affordance

 <p>People could ___ the umbrella.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> stay dry under <input type="checkbox"/> use <input type="checkbox"/> walk under <input type="checkbox"/> twirl 	 <p>People could ___ the remote.</p> <ul style="list-style-type: none"> <input type="checkbox"/> use <input type="checkbox"/> hold <input checked="" type="checkbox"/> play with <input type="checkbox"/> click 	 <p>People could ___ the horse.</p> <ul style="list-style-type: none"> <input type="checkbox"/> saddle <input type="checkbox"/> ped <input checked="" type="checkbox"/> race <input type="checkbox"/> jump 	 <p>People could ___ the cell phone.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> take pictures with <input type="checkbox"/> call with <input type="checkbox"/> text with <input type="checkbox"/> communicate with
---	--	--	---

8. object's position

 <p>The zebra is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> on the wall <input type="checkbox"/> on the right <input type="checkbox"/> on two legs <input type="checkbox"/> behind a barrier 	 <p>The backpack is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> on the man's back <input type="checkbox"/> on the person's shoulder <input type="checkbox"/> on the ground <input type="checkbox"/> with the people 	 <p>The dog is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> standing on two legs <input type="checkbox"/> sitting on a book <input checked="" type="checkbox"/> on the man's lap <input type="checkbox"/> next to the human 	 <p>The pizza is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> in the boy's hands <input type="checkbox"/> on the steering wheel <input type="checkbox"/> in the pizza box <input type="checkbox"/> on the dining table
--	--	---	--

9. person's attribute

 <p>Person B is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> wearing a grey scarf <input type="checkbox"/> a young female child <input type="checkbox"/> a young asian male <input type="checkbox"/> in a white cap 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> a smiling young man in a grey skiing outfit <input type="checkbox"/> a girl in a purple jacket <input checked="" type="checkbox"/> a competitive skier with an orange hat <input type="checkbox"/> wearing a backpack on his back 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> covered with a grey blanket <input type="checkbox"/> a baby with a pacifier <input checked="" type="checkbox"/> female with long dark hair <input type="checkbox"/> wearing sweatpants and athletic shoes 	 <p>Person B is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> a child dressed in black <input type="checkbox"/> a man with dark hair <input type="checkbox"/> a man in a hoodie <input type="checkbox"/> in shorts and a t-shirt
---	---	---	---

10. person's activity

 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> walking across the street <input type="checkbox"/> standing <input type="checkbox"/> ordering food <input type="checkbox"/> selling bananas 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> holding her cellphone <input type="checkbox"/> taking picture of herself <input type="checkbox"/> standing <input type="checkbox"/> walking 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> posing for a photo <input type="checkbox"/> walking <input checked="" type="checkbox"/> surfing <input type="checkbox"/> standing in the water 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> sitting on a chair <input type="checkbox"/> holding an umbrella <input type="checkbox"/> standing <input type="checkbox"/> talking to each other
--	--	---	---

11. person's location

 <p>The people are ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> in a dining area <input type="checkbox"/> in a community room <input type="checkbox"/> at a wedding reception <input type="checkbox"/> in a reception hall 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> at the beach <input type="checkbox"/> on a lake <input type="checkbox"/> at the beach <input type="checkbox"/> in the elevator 	 <p>Person A is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> on the water <input type="checkbox"/> on a lake <input checked="" type="checkbox"/> at the beach <input type="checkbox"/> on a surfboard 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> standing with the group <input checked="" type="checkbox"/> sitting in the snow <input type="checkbox"/> next to the other person <input type="checkbox"/> outside on a mountain
--	---	--	--

12. pair's relationship

 <p>The person is ___ the toilet.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> sitting on <input type="checkbox"/> pouring into <input type="checkbox"/> splitting into <input type="checkbox"/> standing over 	 <p>The people are ___ the teddy bear.</p> <ul style="list-style-type: none"> <input type="checkbox"/> hugging <input type="checkbox"/> holding <input checked="" type="checkbox"/> looking at <input type="checkbox"/> sitting next to 	 <p>The person is ___ the sports ball.</p> <ul style="list-style-type: none"> <input type="checkbox"/> kicking <input type="checkbox"/> holding <input checked="" type="checkbox"/> hitting at <input type="checkbox"/> by the 	 <p>The person is ___ the truck.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> sitting in <input type="checkbox"/> standing near <input type="checkbox"/> looking at <input type="checkbox"/> in line at
---	--	--	--

Figure 4: Examples of correct answers made by nCCA for 12 types of multiple-choice question-answering.

1. image's scene

 <p>This place is a(n) ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> court house <input checked="" type="checkbox"/> bar <input type="checkbox"/> wedding <input type="checkbox"/> street 	 <p>This place is a(n) ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> river <input type="checkbox"/> hacienda <input checked="" type="checkbox"/> vacation spot <input type="checkbox"/> canal 	 <p>This place is a(n) ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> airport <input type="checkbox"/> street <input checked="" type="checkbox"/> subway <input type="checkbox"/> sidewalk 	 <p>This place is a(n) ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> classroom <input type="checkbox"/> living room <input type="checkbox"/> office <input checked="" type="checkbox"/> work place
--	---	--	---

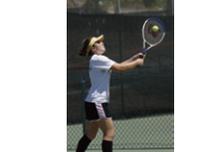
2. image's emotion

 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> hungry <input type="checkbox"/> serious <input type="checkbox"/> silly <input type="checkbox"/> interested 	 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> despair <input type="checkbox"/> concerned <input type="checkbox"/> natural <input type="checkbox"/> healthy 	 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> hungry <input type="checkbox"/> annoyed <input type="checkbox"/> wierd <input type="checkbox"/> guarded 	 <p>When I look at this picture, I feel ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> gentle <input type="checkbox"/> pretty <input type="checkbox"/> tired <input checked="" type="checkbox"/> sad
--	--	--	--

3. image's interesting

 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the guy taking self inside a bathroom <input checked="" type="checkbox"/> the toilet <input type="checkbox"/> girls panties <input type="checkbox"/> the person on the floor 	 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the adorable children <input type="checkbox"/> the injuries <input type="checkbox"/> the bear <input type="checkbox"/> facial expression 	 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the wine <input type="checkbox"/> the people who are celebrating <input checked="" type="checkbox"/> the remaining crist <input type="checkbox"/> the pizza 	 <p>The most interesting aspect of the picture is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the pose of the person <input checked="" type="checkbox"/> frisbee <input type="checkbox"/> the jumping man <input type="checkbox"/> peace sign
---	--	---	--

4. image's past

 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> two men grabbed their phones <input checked="" type="checkbox"/> dad paid for the food <input type="checkbox"/> the knife was down <input type="checkbox"/> she finished eating 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> she picked up the sandwich <input type="checkbox"/> the phone was closed <input type="checkbox"/> the man held a cup <input type="checkbox"/> they labeled the donuts 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the man played tennis <input type="checkbox"/> the woman in red scored a point <input type="checkbox"/> a ball flew by <input type="checkbox"/> he was smiling 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> his phone rang <input type="checkbox"/> people rode motorcycles <input type="checkbox"/> they stopped the scooters <input checked="" type="checkbox"/> the person walked towards the bus
--	---	---	---

5. image's future

 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the two sisters fought over the bear <input checked="" type="checkbox"/> the child hugged the animal <input type="checkbox"/> the girl left for school <input type="checkbox"/> the man made a missy face 	 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the person fell to the ground <input checked="" type="checkbox"/> they will make their first throws <input type="checkbox"/> the person fell on the ground <input checked="" type="checkbox"/> the man is playing with balloon 	 <p>One or two seconds after this picture was taken, ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> the children finished their meal <input type="checkbox"/> the man started to smile <input type="checkbox"/> all will take a sip <input type="checkbox"/> he blew out the candles 	 <p>One or two seconds before this picture was taken, ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> the other person got on the bike <input type="checkbox"/> the guy was still posing for the camera <input type="checkbox"/> the bank transported the bananas down the street <input checked="" type="checkbox"/> the cas was within inches of the motorcycle
---	--	--	--

6. object's attribute

 <p>The tie is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> orange <input checked="" type="checkbox"/> black <input type="checkbox"/> long <input type="checkbox"/> checkered 	 <p>The suitcase is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> black <input type="checkbox"/> rectangle <input type="checkbox"/> big <input type="checkbox"/> colorful 	 <p>The sandwich is ____.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> hot <input type="checkbox"/> a reuben <input type="checkbox"/> partially eaten <input type="checkbox"/> very large 	 <p>The banana is ____.</p> <ul style="list-style-type: none"> <input type="checkbox"/> opened <input type="checkbox"/> green <input type="checkbox"/> picked <input checked="" type="checkbox"/> sliced
---	---	---	---

7. object's affordance

 <p>People could ___ the dining table.</p> <ul style="list-style-type: none"> <input type="checkbox"/> place food on <input checked="" type="checkbox"/> sit on <input type="checkbox"/> put their feet up on <input type="checkbox"/> put things on 	 <p>People could ___ the elephant.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> pet <input type="checkbox"/> care for <input type="checkbox"/> ride <input type="checkbox"/> wash 	 <p>People could ___ the cow.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> herd <input type="checkbox"/> pose by <input type="checkbox"/> race <input type="checkbox"/> use 	 <p>People could ___ the bed.</p> <ul style="list-style-type: none"> <input type="checkbox"/> jump on <input type="checkbox"/> lay on <input type="checkbox"/> sleep <input checked="" type="checkbox"/> sleep on
---	--	---	--

8. object's position

 <p>The chair is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> on the sidelines <input checked="" type="checkbox"/> in the lobby <input type="checkbox"/> behind the boy <input type="checkbox"/> under the girl 	 <p>The teddy bear is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> behind the person <input type="checkbox"/> in his hands <input type="checkbox"/> on the bed <input type="checkbox"/> in the chair 	 <p>The train is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> at the station <input type="checkbox"/> in the station <input type="checkbox"/> on a bridge <input type="checkbox"/> behind the people 	 <p>The sheeps are ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> near the child <input type="checkbox"/> with the guys <input type="checkbox"/> on the grass <input checked="" type="checkbox"/> behind the woman
--	--	--	---

9. person's attribute

 <p>Person A is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> an older guy in red shirt <input checked="" type="checkbox"/> dressed in a fancy coat <input type="checkbox"/> wearing colored jacket with orange stripes <input type="checkbox"/> a small young boy 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> female <input type="checkbox"/> male <input type="checkbox"/> puzzled <input type="checkbox"/> distracted 	 <p>The people are ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> a young girl <input type="checkbox"/> a white male <input type="checkbox"/> children and adults <input type="checkbox"/> wearing red shorts 	 <p>Person A is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> wearing a backpack <input type="checkbox"/> a young woman <input type="checkbox"/> hiding a surfboard <input checked="" type="checkbox"/> dark haried man
--	--	--	---

10. person's activity

 <p>Person A is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> interacting with people on either side <input checked="" type="checkbox"/> looking out of a window <input type="checkbox"/> enjoying the view and sweet moments <input type="checkbox"/> kissing a large stuffed bear 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> sitting <input checked="" type="checkbox"/> eating donut <input type="checkbox"/> displaying donuts <input type="checkbox"/> eating 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> cutting a cake <input type="checkbox"/> eating a sandwich <input type="checkbox"/> thinking <input type="checkbox"/> eating 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> sitting on a chair playing a keyboard <input type="checkbox"/> working on a computer <input type="checkbox"/> holding a phone in her hands <input checked="" type="checkbox"/> typing on the laptop computer
---	---	--	--

11. person's location

 <p>The people are ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> in a large room <input checked="" type="checkbox"/> in a computer lab <input type="checkbox"/> sitting at a table <input type="checkbox"/> sitting by the screen 	 <p>The people are ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> on the street <input type="checkbox"/> at the mall <input type="checkbox"/> on a sidewalk <input type="checkbox"/> in a parade 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> is at a zoo <input type="checkbox"/> at the work farm <input type="checkbox"/> on a dirty road <input type="checkbox"/> in a shallow river 	 <p>The person is ___.</p> <ul style="list-style-type: none"> <input type="checkbox"/> at the kitchen sink <input type="checkbox"/> standing in the doorway <input type="checkbox"/> in the shop <input checked="" type="checkbox"/> in the kitchen
---	--	--	--

12. pair's relationship

 <p>The person is ___ the surfboard.</p> <ul style="list-style-type: none"> <input type="checkbox"/> carrying <input checked="" type="checkbox"/> riding <input type="checkbox"/> walking with <input type="checkbox"/> sitting on 	 <p>The person is ___ the sheeps.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> looking at <input type="checkbox"/> looking after <input type="checkbox"/> taking a picture of <input type="checkbox"/> interacting with 	 <p>The person is ___ the teddy bear.</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> holding <input type="checkbox"/> snuggling <input type="checkbox"/> looking at <input type="checkbox"/> cuddling 	 <p>The person is ___ the sheeps.</p> <ul style="list-style-type: none"> <input type="checkbox"/> enclosing <input type="checkbox"/> petting <input type="checkbox"/> guiding <input checked="" type="checkbox"/> looking at
---	--	---	---

Figure 5: Examples of wrong answers made by nCCA for 12 types of multiple-choice question-answering.

1. image's scene

 <p>nCCA: This place is a(n) <u>video recording</u>. CNN+LSTM: This place is a(n) <u>living room</u>.</p>	 <p>nCCA: This place is a(n) <u>bedroom</u>. CNN+LSTM: This place is a(n) <u>bedroom</u>.</p>	 <p>nCCA: This place is a(n) <u>restaurant</u>. CNN+LSTM: This place is a(n) <u>restaurant</u>.</p>	 <p>nCCA: This place is a(n) <u>parade on a street</u>. CNN+LSTM: This place is a(n) <u>street</u>.</p>
--	--	---	--

2. image's emotion

 <p>nCCA: When I look at this picture, I feel <u>active and excited</u>. CNN+LSTM: When I look at this picture, I feel <u>excited</u>.</p>	 <p>nCCA: When I look at this picture, I feel <u>hungry</u>. CNN+LSTM: When I look at this picture, I feel <u>hungry</u>.</p>	 <p>nCCA: When I look at this picture, I feel <u>cold and free</u>. CNN+LSTM: When I look at this picture, I feel <u>happy</u>.</p>	 <p>nCCA: When I look at this picture, I feel <u>peaceful</u>. CNN+LSTM: When I look at this picture, I feel <u>calm</u>.</p>
---	--	---	--

3. image's interesting

 <p>nCCA: The most interesting aspect of this picture is <u>hat</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the man's face</u>.</p>	 <p>nCCA: The most interesting aspect of this picture is <u>skateboard trick</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the skateboarder</u>.</p>	 <p>nCCA: The most interesting aspect of this picture is <u>the cake</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the food</u>.</p>	 <p>nCCA: The most interesting aspect of this picture is <u>a boy eating donuts</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the man's face</u>.</p>
---	--	---	---

4. image's past

 <p>nCCA: One or two seconds before this picture was taken, <u>he jumped on the skateboard</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the man was on the ground</u>.</p>	 <p>nCCA: One or two seconds before this picture was taken, <u>she sliced the pizza</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the woman was eating</u>.</p>	 <p>nCCA: One or two seconds before this picture was taken, <u>the ball was kicked</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the ball was served</u>.</p>	 <p>nCCA: One or two seconds before this picture was taken, <u>they sliced the pizza</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the woman picked up the pizza</u>.</p>
--	--	---	--

5. image's future

 <p>nCCA: One or two seconds after this picture was taken, <u>the man boarded the bus</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>the man will move forward</u>.</p>	 <p>nCCA: One or two seconds after this picture was taken, <u>the man swallowed his bite</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>the man will eat the food</u>.</p>	 <p>nCCA: One or two seconds after this picture was taken, <u>he went skateboarding</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>the man will go down the slope</u>.</p>	 <p>nCCA: One or two seconds after this picture was taken, <u>the skier landed</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>he fell</u>.</p>
--	---	---	---

6. object's attribute

 <p>nCCA: The boat is <u>filled with fruit</u>. CNN+LSTM: The boat is <u>wooden</u>.</p>	 <p>nCCA: The sports ball is <u>white and orange</u>. CNN+LSTM: The sports ball is <u>white</u>.</p>	 <p>nCCA: The teddy bear is <u>grey</u>. CNN+LSTM: The teddy bear is <u>brown</u>.</p>	 <p>nCCA: The bed is <u>blue</u>. CNN+LSTM: The bed is <u>covered with a blanket</u>.</p>
---	---	---	--

1. image's scene

 <p>nCCA: This place is a(n) <u>video recording</u>. CNN+LSTM: This place is a(n) <u>living room</u>.</p>	 <p>nCCA: This place is a(n) <u>bedroom</u>. CNN+LSTM: This place is a(n) <u>bedroom</u>.</p>	 <p>nCCA: This place is a(n) <u>restaurant</u>. CNN+LSTM: This place is a(n) <u>restaurant</u>.</p>	 <p>nCCA: This place is a(n) <u>parade on a street</u>. CNN+LSTM: This place is a(n) <u>street</u>.</p>
--	--	---	--

2. image's emotion

 <p>nCCA: When I look at this picture, I feel <u>active and excited</u>. CNN+LSTM: When I look at this picture, I feel <u>excited</u>.</p>	 <p>nCCA: When I look at this picture, I feel <u>hungry</u>. CNN+LSTM: When I look at this picture, I feel <u>hungry</u>.</p>	 <p>nCCA: When I look at this picture, I feel <u>cold and free</u>. CNN+LSTM: When I look at this picture, I feel <u>happy</u>.</p>	 <p>nCCA: When I look at this picture, I feel <u>peaceful</u>. CNN+LSTM: When I look at this picture, I feel <u>calm</u>.</p>
---	--	---	--

3. image's interesting

 <p>nCCA: The most interesting aspect of this picture is <u>hat</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the man's face</u>.</p>	 <p>nCCA: The most interesting aspect of this picture is <u>skateboard trick</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the skateboarder</u>.</p>	 <p>nCCA: The most interesting aspect of this picture is <u>the cake</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the food</u>.</p>	 <p>nCCA: The most interesting aspect of this picture is <u>a boy eating donuts</u>. CNN+LSTM: The most interesting aspect of this picture is <u>the man's face</u>.</p>
---	--	---	---

4. image's past

 <p>nCCA: One or two seconds before this picture was taken, <u>he jumped on the skateboard</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the man was on the ground</u>.</p>	 <p>nCCA: One or two seconds before this picture was taken, <u>she sliced pizza</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the woman was eating</u>.</p>	 <p>nCCA: One or two seconds before this picture was taken, <u>the ball was kicked</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the ball was served</u>.</p>	 <p>nCCA: One or two seconds before this picture was taken, <u>they sliced the pizza</u>. CNN+LSTM: One or two seconds before this picture was taken, <u>the woman picked up the pizza</u>.</p>
--	--	---	--

5. image's future

 <p>nCCA: One or two seconds after this picture was taken, <u>the man boarded the bus</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>the man will move forward</u>.</p>	 <p>nCCA: One or two seconds after this picture was taken, <u>the man swallowed his bite</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>the man will eat the food</u>.</p>	 <p>nCCA: One or two seconds after this picture was taken, <u>he went skateboarding</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>the man will go down the slope</u>.</p>	 <p>nCCA: One or two seconds after this picture was taken, <u>the skier landed</u>. CNN+LSTM: One or two seconds after this picture was taken, <u>he fell</u>.</p>
--	---	---	---

6. object's attribute

 <p>nCCA: The boat is <u>filled with fruit</u>. CNN+LSTM: The boat is <u>wooden</u>.</p>	 <p>nCCA: The sports ball is <u>white and orange</u>. CNN+LSTM: The sports ball is <u>white</u>.</p>	 <p>nCCA: The teddy bear is <u>grey</u>. CNN+LSTM: The teddy bear is <u>brown</u>.</p>	 <p>nCCA: The bed is <u>blue</u>. CNN+LSTM: The bed is <u>covered with a blanket</u>.</p>
---	---	--	--

Figure 6: Examples of focused sentence generation achieved by nCCA and CNN+LSTM.