







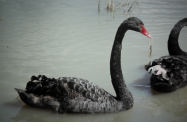

Image+Caption	Question	Human Answer	CoAtt-G-MLE	Ours
 <p>A long bridge with cars has a train going underneath</p>	<p>Can you see people? Inside or outside? Can you see cars? Car, truck, or minivan? Is the bridge over water? What is it over? What the bridge is made out of? Can you see trains in the yard? How many cars? All the same?</p>	<p>No i can't It is outside I see 1 It seems to be a small car No it is not It is over a train yard The road appears to be concrete or asphalt Yes i can It has 3 The first is an engine</p>	<p>No Outside No i can't I can't tell Yes it is I can't tell Wood No i can not I see 4 Yes they are</p>	<p>No It is outside I can see 1 I think it's a car No it is not I can't tell It looks like concrete No i can not I can see 3 Yes they are</p>
 <p>Surfers are practicing in the shore of a beach</p>	<p>How many surfers are there? How hard are they practicing? What 's the weather like? What gender do you think they identify as? How old would you guess they are? Does it look like 1 is teaching the other? What are the water conditions like? Would you guess it 's a good day for surfing? What type of beach are they on? Are the wearing professional gear?</p>	<p>2 are practicing Not very hard Cloudy but sunny 1 female and 1 male 20s Not really Calm Yes Sandy public beach No</p>	<p>5 No Sunny Male Late teens Yes Water Yes Beach No</p>	<p>I see 5 Not very Mostly sunny 1 male and female 20s Yes Water Yes Sandy beach No</p>
 <p>A bear walking through a field next to a forest</p>	<p>What color is the grass? Are there any people in the image? Is the bear grown or a cub? Are there any artifacts in the image? How big does the field appear to be? Are there any flowers growing in the field? What time of day does it appear to be? Is the bear facing the camera? Are there any trees nearby? How many trees are there?</p>	<p>Some green and some yellow No Looks fairly young, maybe a few years No Quite large No Maybe close to dark in a few hours No Yes Many all grown together</p>	<p>Green No Grown No Big No Afternoon No Yes Many</p>	<p>Mostly green No Looks young No It looks like a large hill No Afternoon No Yes Many trees</p>
 <p>A man standing next to a parked motorcycle on a dirt road</p>	<p>Is there a harley motorcycle? What makes are the motorcycles? Are they dirty? What color are they? Do they have any type of cover on them? Are UNK used to attach the luggage? What color is the luggage? What color are the UNK? How many pieces of luggage are on the bikes? Can you see the handles?</p>	<p>No Not sure more like dirt bikes A little Black No a lot of luggage though Yes Orange on one , gray on the other White 1 UNK 1 on the back of each Yes</p>	<p>No They are harley No Red Yes No Black Black 3 Yes</p>	<p>No Not sure Yes, a little Red and black No Yes Orange and gray Black 3 Yes</p>
 <p>A black and white fighter plane flying over a grassy area</p>	<p>Is it a passenger plane? Is there a logo on the plane? Can you tell what airline the plane is from? What color is the plane? Can you see the control tower? Is the landing gear down? How far off the ground does the plane look? Is it sunny? Does it look like it might rain? Do you see a fence near the runway?</p>	<p>Yes There is 'UNK' on the side No, the plane is too far away It 's navy blue with light blue No Yes Hard to say, 100 ft maybe Very cloudy Maybe the clouds aren't dark I see a fence</p>	<p>No Yes No Grey No No Hard to say No No No</p>	<p>Yes Yes there is one No Grey with a blue stripe No Yes It 's a very high up, but hard to tell No, it 's cloudy No I can see a fence</p>
 <p>One pizza is fully eaten while other pizzas still have slices left</p>	<p>What is the color of the tablecloth? How many people are seated at the table? Is the person wearing any rings? How many dishes are there on the table? How many plates are there for eating? No sorry and apologies, are there any serving cutlery? Are the fingers visible are that of a lady? What is the size of the pizza? What kind of pizza it is? Any drinks on the table?</p>	<p>Black and white I can only see the fingers of 1 Not that i can see 4 Correction: there are 2 plates 1 fork, 2 knives Hard to tell probably not There was a lot of pizza and more I can't see any toppings looks like 1 glass of coke</p>	<p>Red and white 0 No 2 2 No No Small Cheese and pepperoni Yes</p>	<p>Black and white 0 No, i can't see 2 2 Fork and knives No It 's a slice size Cheese and pepperoni Coke</p>
 <p>A black swan has some wet grasses in its mouth</p>	<p>Are the geese grown? Is this a lake? What color is the water? Are there any other birds? Are the birds the same size? Are there any people? Do you see any trees? Is it daytime? Is it sunny outside? Are the geese making ripples?</p>	<p>Yes Yes it is a lake or a pond The water is brown No just the 2 geese They are No there are 0 No i can not see any Yes it is Yes, it is sunny Not really</p>	<p>Yes I can't tell The water is blue No Yes they are No people No I can't tell I can't tell No</p>	<p>Yes Yes, i think so The water is clear No, just the 1 geese Yes they are No, there are no people No i do not Yes it is Yes it is No they are not</p>
 <p>A pizza that is sitting on a wooden board</p>	<p>Is it a cutting board? Is it large or small? What size pizza? What veggies? Is it sliced? Is it bell peppers? Are the carrots cut? Any other food? What kind of crust? Does it look old or fresh?</p>	<p>Yes It 's a larger square board Maybe a medium size Tomato, carrots maybe, cucumbers No it 's whole I 'm not sure , just smaller pieces Yes , I guess they are carrots No It 's homemade pizza Fresh</p>	<p>No Large Large Spinach No No, it 's green Yes No No Fresh</p>	<p>No Large It's a large pizza I see spinach and carrots No No, it 's a basil Yes No It's thin crust It looks fresh</p>

Figure 1: Qualitative results of our model (CoAtt-GAN-w/ R_{inte} -TF) comparing to human ground-truth answer and our baseline model.