Appendix

Vishvak Murahari[1] **Prithvijit Chattopadhyay**[1]
Dhruv Batra[1,2] **Devi Parikh**[1,2] **Abhishek Das**[1]

[1]Georgia Tech [2]Facebook AI Research

1 Qualitative Output

The dialog generated by self-talk between different SL model variants is provided in Table. 3 and different RL model variants is provided in Table. 4. We observe the variants with Diverse-Q-BOT tend to generate more diverse, image relevant and fluent dialog.

2 Experiments with Diverse-A-BOT

We also ran experiments where we used the repetition penalty (Eqn. (1) in the main paper) during SL pre-training of A-BOT. In Tables 1 and 2 we report results on the A-BOT retrieval metrics (Das et al., 2017a) and the diversity metrics respectively. In Fig. 1, we provide the performance of various Q-BOT variants paired with this Diverse-A-BOT on the image-guessing task.

We note that retrieval metrics for Diverse-A-BOT are better than vanilla SL: A-BOT. Therefore, this repetition penalty does help significantly during supervised pre-training. However, finetuning the Diverse-A-BOT via RL does not lead to significant improvements.

We note that for the diversity metrics on Q-BOT-A-BOT dialog, self-talk with Diverse-A-BOT does not change the diversity metrics significantly. We observe the same trend in image-guessing performance as well.

3 Model Architecture details

We use a Hierarchical Recurrent Encoder (HRE) for representing dialog context. In this encoder the image representation is concatenated with every question word when fed to the LSTM. We then encode each QA-pair in the dialog history with another LSTM with shared weights. The image-question representation, computed for every round from 1 through t, is concatenated with history representation from the previous round. This gives us a set of t question-history vectors for t rounds. These vectors are fed as input to a dialog-level LSTM, whose output state at t is used to decode the response to Q_t. 2 shows the model architecture of the HRE.

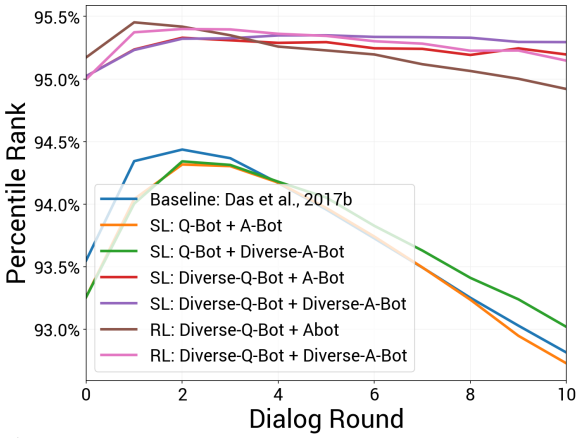


Figure 1: Performance on the image-guessing task. Percentile rank (higher is better) of the true image (shown to A-BOT) as retrieved using fc7 image feature predictions from Q-BOT.

All LSTMs are 2-layered with 512-dim hidden states. We learn 300-dim embeddings for words and images. These word embeddings are shared across question, history, and decoder LSTMs.

References

Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, and Dhruv Batra. 2017a. Visual Dialog. In *CVPR*. 1, 2, 3, 4

Abhishek Das, Satwik Kottur, José M.F. Moura, Stefan Lee, and Dhruv Batra. 2017b. Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning. In *ICCV*. 2, 4

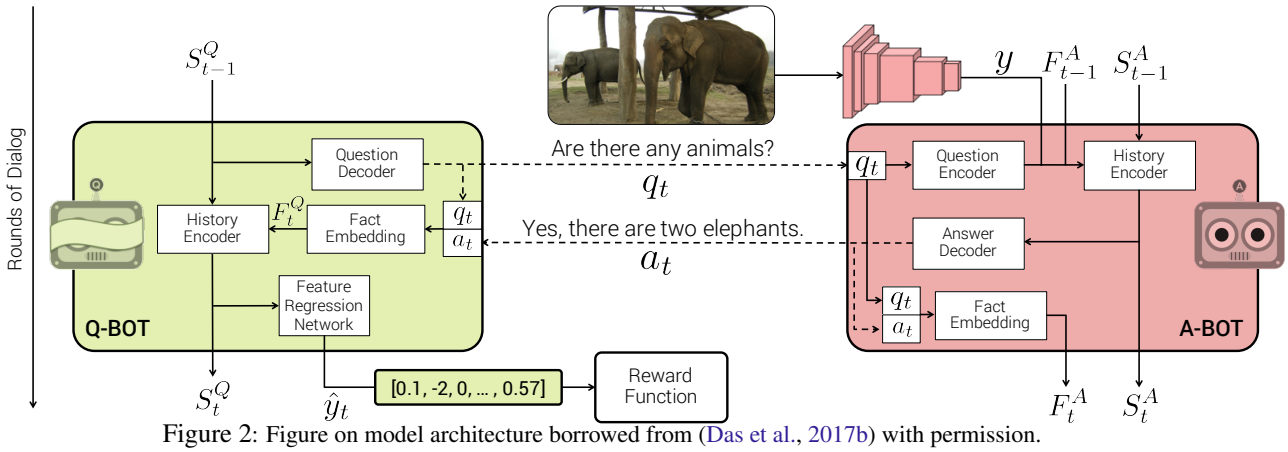


Figure 2: Figure on model architecture borrowed from (Das et al., 2017b) with permission.

	v1.0 val						v1.0 test-std					
	NDCG ↑	MRR ↑	R@1 ↑	R@5 ↑	R@10 ↑	Mean Rank ↓	NDCG ↑	MRR ↑	R@1 ↑	R@5 ↑	R@10 ↑	Mean Rank ↓
Baseline: Das et al. (2017b)	53.76	46.35	36.22	56.15	62.41	19.34	51.60	45.67	35.05	56.30	63.25	19.15
Diverse SL A-BOT	53.82	46.55	36.46	56.16	62.68	19.34	51.77	45.98	35.58	56.33	63.08	18.97
Diverse RL A-BOT (Finetuned with Diverse-A-BOT)	53.94	46.67	36.54	56.34	62.99	19.24	51.80	45.70	34.85	56.33	63.90	18.88

Table 1: A-BOT performance on the VisDial v1.0 dataset (Das et al., 2017a). Arrow indicates direction of better performance.

	Diversity							Relevance
	# Novel questions ↑	# Unique questions ↑	Mutual overlap ↓	Ent-1 ↑	Ent-2 ↑	Dist-1 ↑	Dist-2 ↑	Negative log likelihood ↓
Baseline: Das et al. (2017b)	71	6.70 ± 0.07	0.58 ± 0.01	2.72 ± 0.01	3.03 ± 0.02	0.35 ± 0.0	0.43 ± 0.0	9.94
SL: Q-BOT + Diverse-A-BOT	53	6.58 ± 0.07	0.60 ± 0.01	2.71 ± 0.01	3.00 ± 0.01	0.34 ± 0.0	0.42 ± 0.0	10.05
SL: Diverse-Q-BOT + Diverse-A-BOT	136	7.42 ± 0.07	0.50 ± 0.01	2.81 ± 0.01	3.18 ± 0.01	0.38 ± 0.0	0.48 ± 0.0	10.10
RL: Diverse-Q-BOT + Diverse-A-BOT	288	7.87 ± 0.06	0.46 ± 0.01	2.86 ± 0.01	3.24 ± 0.01	0.39 ± 0.0	0.51 ± 0.0	10.32

Table 2: Q-BOT diversity and relevance on v1.0 val. ↑ indicates higher is better. ↓ indicates lower is better.

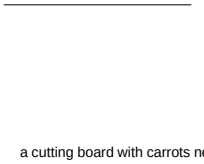
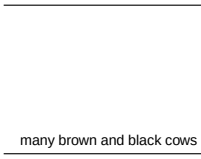
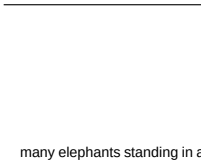
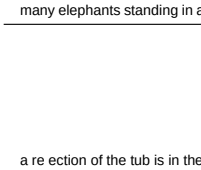
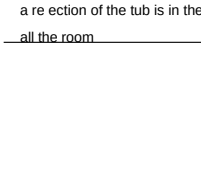
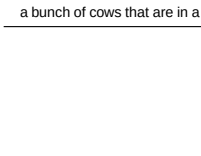
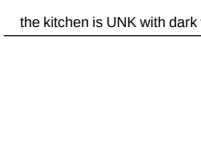
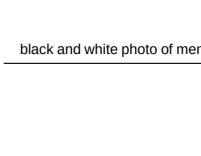
Image + Caption	Human-Human dialog (Das et al., 2017a)	Q-BOT-A-BOT dialog	Diverse-Q-BOT-A-BOT dialog
 a cutting board with carrots next to a cell phone.	what color is the cutting board? blue how big is it about the size of a regular laptop screen how many carrots are there ? 4 are they the orange kind ? 2 are orange , 2 are white what kind of cell phone is it ? a samsung can you see any text on the screen ? a recipe for UNK UNK soup do you see the room ? just the counter are there people ? not visible what color is the counter ? white	is the photo in color ? no how many people do you see ? no people this a close up ? no is this a close up ? zebra is is this a close up ? no is this a close up ? no is this a close up ? no is this a close up ? no is this a close up ? no is this a close up ? no	is the photo in color ? 1 picture is in color is the photo close up ? black and white is the photo in color ? no it 's black and white is it a close up ? no it 's a close up of the cake can you tell what brand it is ? no i ca n't see the brand can you tell if it 's a male or female ? no i ca n't tell can you tell if it 's a male or female ? no i ca n't tell is this in a kitchen ? nope it 's in a home are there any utensils ? no are there any utensils ? nope
 many brown and black cows standing in the grass	is the eld large ? yes is the grass green or brown ? green can you see a barn ? no can you see a fence ? no can you see a water supply ? yes can you see any people ? no can you see trees ? yes can you see mountains in the background ? no is the grass tall or short ? short	what color are the cows ? they are brown and white how many cows are there ? 4 what color are the cows ? they are brown and white can you see the sky ? it is not visible do you see any trees ? i do not see any trees can you see the sky ? green sky are there any clouds ? the sky is not visible how many cows are there ? 2 what color are the cows ? the cows are brown and white what color are the cows ? they are brown	do you see any people ? 1 do you see a fence ? a small 1 how many cows are there ? 5 what color are the cows ? black and white can you see the sky ? some can you see the sky ? nope are there any other animals ? not that i can see can you see the sky ? not at all are there any trees ? in the background can you see the sky ? not at all
 many elephants standing in an enclosure near a building	what type of building ? may be part of a zoo how many elephants ? about 4 is it sunny ? clear day , yes any trees ? yes what color is the building ? dark grey are there other animals ? no are there any people ? yes what are they wearing ? shorts and shirts how many ? 3	how many elephants are there ? ca n't tell what color are the elephants ? ca n't tell how many elephants are there ? ca n't tell what color is the fence ? i ca n't tell can you see the sky ? i ca n't see the sky are there any trees ? just a few does this look like a zoo ? 1 of them is are there any people ? nothing in the picture how many elephants are there ? just 1 can you see the sky ? just a little	how many elephants are there ? 1 any people ? ca n't see any male or female ? not sure what is the elephant doing ? just standing there male or female ? there are 2 males and 1 female can you see the sky ? 1 of it what color is the building ? just a black and white picture do you see any trees ? there are trees in the background are there any other animals ? red and white do you see any trees ? i see some in the background
 a re ection of the tub is in the bathroom mirror along with the sink , you can see all the room	is the photo in color ? yes is there humans in the photo ? no is there any animals in the photo ? no what color are the walls ? tan what color is the sink ? white is there any towels ? no is there any pictures on the wall ? no what color is the tub ? white can you see the oor ? no	is the photo in color ? 5 of them are are there any people ? it 's a close up so i ca n't see any people what color is the sink ? about the same color as the toilet do you see a mirror ? no what color is the sink ? yes it is white what color is the sink ? yes it is white what shape is the mirror ? yes , it 's a rectangle are there any towels ? he 's on a towel rack are there any towels ? i do n't see any towels are there any towels ? i do n't see any towels	is the photo in color ? no is there any people ? 5 what color walls ? i ca n't see the walls is there a mirror ? yes is there a mirror ? yes is there a mirror ? yes are the walls seen ? no what color are the walls ? no walls what is the oor made of ? yes what color is the oor ? i ca n't see the oor
 a bunch of cows that are in a grassy eld	about how many cows ? over a dozen are there any calves ? no are they eating ? yes is it sunny ? no is there a fence around them ? no do you see any buildings ? no mountains in the background ? no is there snow on the ground ? no any trees ? yes	is the photo in color ? yes is there any people ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no	how many cows are there ? about 10 what color are the cows ? black and white how many cows are there ? about 10 what color are the cows ? black and white what is the weather like ? sunny are there any trees ? no are there any people ? no are there any trees ? no are there any buildings ? no are there any trees ? yes
 the kitchen is UNK with dark wood cabinets and a nearby staircase	what color are the kitchen walls ? white is there a window present ? yes are there pictures paintings hanging up ? no is the stove visible ? yes what color is the stove ? black and silver are there any people visible ? no what color is the nearby staircase ? dark brown is the staircase carpeted ? no are the handrails of the staircase visible ? yes	is the photo in color ? yes is there any people ? no is there a stove ? no are there any windows ? no can you see the oor ? no can you see the oor ? no can you see the oor ? yes can you see the oor ? yes can you see the oor ? yes are the walls visible ? yes	is the kitchen clean ? brown is the kitchen clean ? it 's cluttered are there any windows ? yes is it daytime ? yes are there any people ? yes are the appliances stainless steel ? i ca n't tell are the people in the kitchen ? yes is the kitchen clean ? yes are the people male or female ? yes is the kitchen clean ? yes
 black and white photo of men working in a kitchen	how many men ? 3 how big is the kitchen ? small industrial what is the age range ? teens to 40s is the kitchen belong in a house or a restaurant ? restaurant what are they doing in the kitchen ? cooking and mixing food what kind of oor in the kitchen ? i ca n't tell is there any decorations in the kitchen ? no is the kitchen look clean ? no does the men seem to be UNK ? they seem UNK	can you see the fridge ? the side of it what color is it ? the photo is black and white can you see the sink ? 1 of it can you see the sink ? just a tiny bit of it can you see the sink ? 0 can you see the sink ? does n't look like it can you see the sink ? 0 can you see the sink ? 4 of them can you see the fridge ? UNK can you see the fridge ? not real	how many men are there ? looks like 2 men and 1 woman how old do they look ? it 's hard to tell , but they look like they are in their 20s how old do they look ? looks to be in their 30 's are they wearing aprons ? just 1 of them is wearing hat are they wearing hats ? it is hard to tell , but i think so can you see the stove ? not at all are there any pots on the stove ? not that i can see does the kitchen look clean ? black and white photo but i think it 's clean can you see any windows ? no do you see any windows ? ca n't see any windows
 a half-eaten chicken breast is next to an UNK piece of bread and a small serving of green beans and carrots	who is eating ? is the drink in the cup soda what is the food on ? bread and vegetable is it in a plate ? what kind of sauces are there in the back i am asking the questions here ? yes you are what is the food on ? bread , vegetables , chicken and drink are they on a table ? yes they are any people ? there are n't any people in sight what color is the table ? it is dark gray any cutlery ? yes there is a butter knife in the plate	is the photo in color ? it is is the photo close up ? no is the photo in color ? it is do you see any other people ? it 's a close up of the man 's hand , so i ca n't tell how old is the man ? maybe in his 30s is he wearing glasses ? i ca n't see his face what color is his shirt ? not sure , the photo is in black and white does he have facial hair ? no does he wear glasses ? yes	are there any people in the photo ? yes is the photo in color ? it is is it a professional photo ? i do n't think so how many people ? 0 is there a tablecloth on the table ? yes what color is the tablecloth ? white are there any utensils visible ? no what color is the plate ? it 's white are there any utensils visible ? no are there any utensils visible ? no

Table 3 Selected examples of dialog on 1.0 VisDial test split for different SL variants. We observe that variant involving Diverse-Q-BOT generates more diverse, image relevant and uent dialog.

Image + Caption	Human-Human dialog (Das et al., 2017a)	(Das et al., 2017b) dialog	RL: Diverse-Q-BOT-A-BOT dialog
 a cutting board with carrots next to a cell phone.	what color is the cutting board? blue how big is it about the size of a regular laptop screen how many carrots are there ? 4 are they the orange kind ? 2 are orange , 2 are white what kind of cell phone is it ? a samsung can you see any text on the screen ? a recipe for UNK UNK soup do you see the room ? just the counter are there people ? not visible what color is the counter ? white	is the photo in color ? no is the photo close up ? no how many people ? cows is this a close up ? 2 of them are what are they wearing ? no shirt and jeans what are they wearing ? no shirt and jeans do you see other people ? sunny 1 in the background what color are the scissors ? no handle is this in a kitchen ? yes do you see the counter ? no	is the photo in color ? black and white is there any people ? 1 hand is this indoors or outdoors ? no idea is the person male or female ? up close , i ca n't tell can you tell what kind of sandwich it is ? no is there a knife ? no is there a cutting board ? yes are the carrots cut up ? daytime is the scene well lit ? no are the carrots raw or cooked ? short
 many brown and black cows standing in the grass	is the field large ? yes is the grass green or brown ? green can you see a barn ? no can you see a fence ? no can you see a water supply ? yes can you see any people ? no can you see trees ? yes can you see mountains in the background ? no is the grass tall or short ? short	can you see any people ? nope can you see the sky ? a little bit are there any clouds ? i ca n't see the sky are there any other animals ? not that i can see do you see any trees ? lots of trees how many cows are there ? ca n't tell , they are far away do you see any buildings ? i do not can you see the sky ? ca n't see the sky do you see any clouds ? ca n't see the sky can you see the sky ? ca n't see the sky	are the cows all the same color ? they are different shades of brown do any of the cows have horns ? it does n't look like it do you see a fence ? there is no fence how many cows are there ? 4 can you see the sky ? 1 of it what color is the grass ? the grass is green can you see the sky ? it is not visible does it appear to be sunny ? not really can you see the sky ? not at all can you see the ground ? a little bit
many elephants standing in an enclosure near a building	what type of building ? may be part of a zoo how many elephants ? about 4 is it sunny ? clear day , yes any trees ? yes what color is the building ? dark grey are there other animals ? no are there any people ? yes what are they wearing ? shorts and shirts how many ? 3	do you see people ? 2 how many elephants ? not sure does the elephant have tusks ? not that i can see does the elephant have tusks ? in the distance , yes does the elephant have tusks ? not that i can see any other animals ? not that i can see does the elephant have tusks ? not that i can see does the elephant have tusks ? not that i can see does the elephant have tusks ? not that i can see can you see the sky ? 1 of it	can you see any people ? yes what are the elephants doing ? just standing around how many people are there ? a lot , too many to count does this look like a zoo ? not really how many elephants are there ? in the distance there are many does this look like a parade ? a little bit does this look like a circus ? nope can you see any signs ? there is a sign on the street does it look like it is in the usa ? i do n't think so can you see the sky ? i can
a reflection of the tub is in the bathroom mirror along with the sink , you can see all the room	is the photo in color ? yes is there humans in the photo ? no is there any animals in the photo ? no what color are the walls ? tan what color is the sink ? white is there any towels ? no is there any pictures on the wall ? no what color is the tub ? white can you see the floor ? no	how big is the bathroom ? ca n't see the whole room how big is the mirror ? i can only see a small part of it what color is the sink ? there is no sink what color is the sink ? black what color is the sink ? yes what color are the walls ? he is white is there a mirror ? yes are there any people ? he is alone are there any animals ? white cat what color is the sink ? yes	how big is the bathroom ? it 's pretty big is it clean ? maybe how big is the mirror ? i can only see a small part of it what color is the wall ? maybe white is there a shower curtain ? there is no shower curtain is there a mirror ? yes what shape is the mirror ? yes , it 's a rectangle what color is the wall ? i ca n't see the wall are there any people ? he 's the only 1 is there a window ? i do n't see 1
a bunch of cows that are in a grassy field	about how many cows ? over a dozen are there any calves ? no are they eating ? yes is it sunny ? no is there a fence around them ? no do you see any buildings ? no mountains in the background ? no is there snow on the ground ? no any trees ? yes	is the photo in color ? yes is there any people ? no how many cows ? black and white cows is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no is there a fence ? no	how many cows are there ? about 10 what color are the cows ? black and white are any of them babies ? no how many cows are there ? about 10 are there any people in the photo ? no are there any other animals in the photo ? no are there any buildings in the photo ? no are there any trees ? yes is there a fence ? no is it sunny out ? yes
the kitchen is UNK with dark wood cabinets and a nearby staircase	what color are the kitchen walls ? white is there a window present ? yes are there pictures paintings hanging up ? no is the stove visible ? yes what color is the stove ? black and silver are there any people visible ? no what color is the nearby staircase ? dark brown is the staircase carpeted ? no are the handrails of the staircase visible ? yes	is this in color ? no do you see any people ? no do you see a stove ? standing are there any windows ? white windows can you see the floor ? metal floor can you see the floor ? brown tile can you see the floor ? yes is it tiled ? no do you see a window ? short window are there curtains on the window ? yes	what color are the appliances ? black and white is there a window ? yes is there a sink ? yes can you see a refrigerator ? yes are there any people ? yes is it daytime ? nope is it sunny out ? just a little is there a window ? brown can you see the floor ? no do you see a stove ? no
black and white photo of men working in a kitchen	how many men ? 3 how big is the kitchen ? small industrial what is the age range ? teens to 40s is the kitchen belong in a house or a restaurant ? restaurant what are they doing in the kitchen ? cooking and mixing food what kind of floor in the kitchen ? i ca n't tell is there any decorations in the kitchen ? no is the kitchen look clean ? no does the men seem to be UNK ? they seem UNK	can you see the fridge ? yep can you see the fridge ? kind of does it have a freezer on it ? nope how many men are there ? ca n't tell how old do they look ? it 's hard to tell , but i 'd say in their 30 's what color are the walls ? it 's a black and white photo so i ca n't tell how old are the men ? i ca n't see their faces , i do n't know does the kitchen look clean ? 0 how old do the men look ? i ca n't see their faces , i do n't know are they wearing aprons ? a few are	can you tell what kind of food they are ? i can not tell what kind of food it is does the kitchen look clean ? fairly clean do you see any windows ? partially in the background do you see any curtains ? maybe 1 in the background is it day or night ? ca n't tell are the men old or young ? ca n't see their faces are they wearing hats ? just 1 are they wearing glasses ? it looks like they are are they smiling ? i ca n't see their faces are they smiling ? i ca n't see their faces
a half-eaten chicken breast is next to an UNK piece of bread and a small serving of green beans and carrots	who is eating ? is the drink in the cup soda what is the food on ? bread and vegetable is it in a plate ? what kind of sauces are there in the back i am asking the questions here ? yes you are what is the food on ? bread , vegetables , chicken and drink are they on a table ? yes they are any people ? there are n't any people in sight what color is the table ? it is dark gray any cutlery ? yes there is a butter knife in the plate	are there any people ? i see part of a person 's hand what color is the plate ? black and white are there any utensils ? no are there any utensils ? no are there any utensils ? silver spoon are there any utensils ? beige spoon are there any utensils ? i see a fork are there any utensils ? yes a fork and a knife are there any napkins ? 0 that i can see are there any napkins ? it 's a close up so i ca n't tell	are there any people in the picture ? yes are they inside or outside ? outside what gender are the people ? i ca n't tell what are the people doing ? 1 is looking at the camera , the other is looking at the camera can you tell if they are male or female ? i ca n't tell can you tell if the person is male or female ? i ca n't see their face , i do n't know can you tell if they are male or female ? i ca n't tell can you tell what kind of sandwich it is ? it has some meat on it , but i ca n't tell what it is is it on a plate ? no what color is the handle of the knife ? silver

Table 4: Selected examples of dialog on v1:0 VisDial test split for different RL variants. We observe that variant involving Diverse-Q-BOT generates more diverse, image relevant and fluent dialog.