Akshat Agarwal
Swaminathan G.
Vasu Sharma
Katia Sycara

Carnegie Mellon University
School of Computer Science

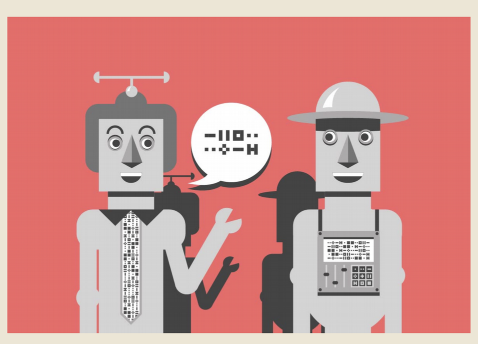
2018 CVPR
SALT LAKE CITY • JUNE 18-22

Mind Your Language: Learning Visually Grounded Dialog in a Multi-Agent Setting

Abstract

- Humans adhere to natural language because they have to interact with an entire community
- Having a private language for each person would be inefficient

We propose a multi-agent dialog framework (MADF) where each agent interacts with and learns from multiple agents and show that it results in more coherent and **human-interpretable dialog** between agents, without compromising on task performance



Interpretable, goal-oriented dialog between artificial agents

Problem Statement

- Formulated as a conversation between two collaborative agents, a Question (Q-) Bot and an Answer (A-) Bot
- A-Bot given an image, while Q-Bot is given only a caption to the image - both agents share a common objective, which is for Q-Bot to form an accurate mental representation of the unseen image
- Facilitated by exchange of 10 pairs of questions and answers between the two agents, using a shared common vocabulary
- Pretraining the agents with supervision from the VisDial dataset, followed by making them interact and adapt to each other via reinforcement learning maximizes task performance, but the agents learn to communicate in non-grammatical and semantically meaningless sentences, hence motivating our multi-agent setup

Method

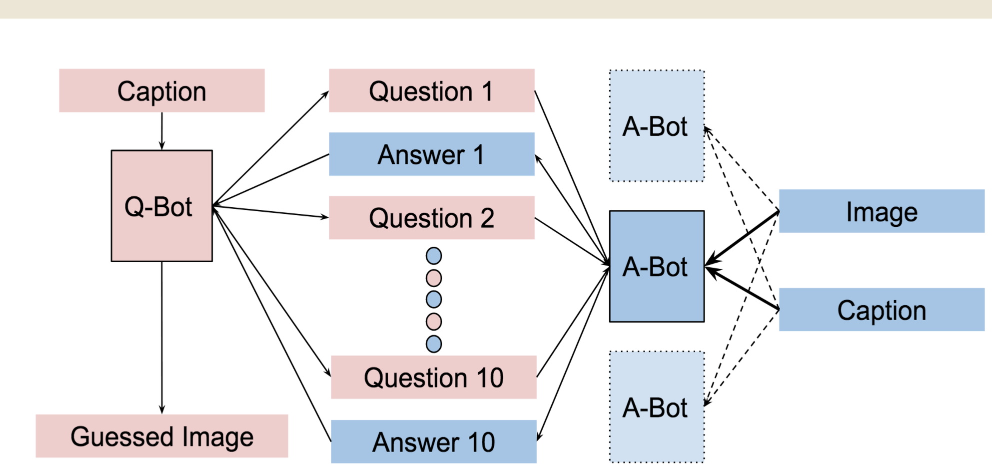
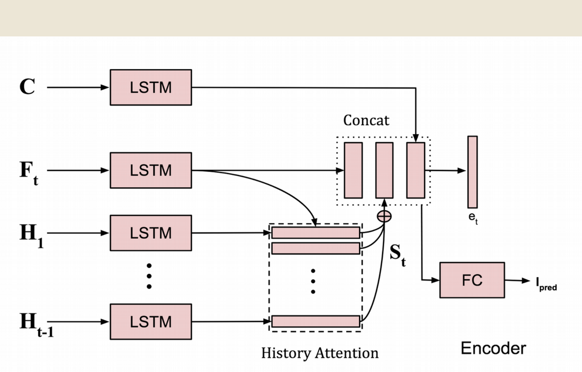
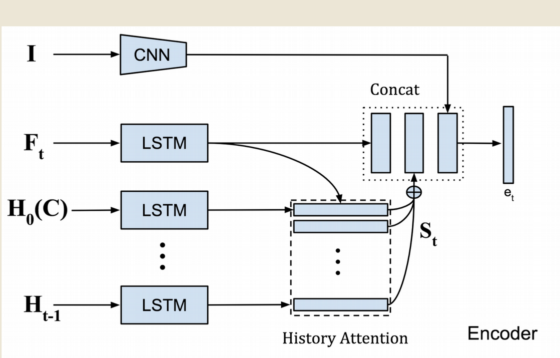


Figure 1: Multi-Agent (with 1 Q-Bot, 3 A-Bots) Dialog Framework



(a) Encoder architecture for Q-Bot



(b) Encoder architecture for A-Bot

- 1 Q-Bot and 1 A-Bot are trained in isolation via supervision to optimize the MLE objective, leading to uninformative and repetitive dialog and inability to respond to out of distribution questions/answers
- We use Curriculum based learning to smoothly transition from supervised learning to Reinforcement Learning

- Reinforcement Learning uses the change in distance between the predicted image embedding and the ground truth embedding as the reward function which is shared by both the Q and A bot. We train the system using the REINFORCE algorithm.

$$r_t(s_t^Q, (q_t, a_t, y_t)) = l(\hat{y}_{t-1}, y^{gt}) - l(\hat{y}_t, y^{gt})$$

- No explicit incentive to maintain natural language and hence prone to deviate from it to optimize transfer of information between bots
- We solve the problem using our Multi Agent setup where we arbitrarily pick a Q and A bot pair and carry out a round of training for them and keep repeating the process
- Much harder for the bots to deviate from natural language in this setting as coming up with a new language pair for each pair of bots is highly inefficient

Algorithm 1 Multi-Agent Dialog Framework (MADF)

```
procedure MultiBotTrain
    while train_iter < max_train_iter do                         ▷ Main Training loop over batches
        Qbot ← random_select (Q_1, Q_2, Q_3....Q_q)
        Abot ← random_select (A_1, A_2, A_3....A_a)              ▷ Either q or a is equal to 1
        history ← (0, 0, ...0)                                    ▷ History initialized with zeros
        fact ← (0, 0, ...0)                                       ▷ Fact encoding initialized with zeros
        Δimage_pred ← 0                                           ▷ Tracks change in Image Embedding
        Qz_1 ← Ques_enc(Qbot, fact, history, caption)
        for t in 1:10 do                                          ▷ Have 10 rounds of dialog
            ques_t ← Ques_gen(Qbot, Qz_t)
            Az_t ← Ans_enc(Abot, fact, history, image, ques_t, caption)
            ans_t ← Ans_gen(Abot, Az_t)
            fact ← [ques_t, ans_t]                                ▷ Fact encoder stores the last dialog pair
            history ← concat(history, fact)                       ▷ History stores all previous dialog pairs
            Qz_t ← Ques_enc(Qbot, fact, history, caption)
            image_pred ← image_regress(Qbot, fact, history, caption)
            R_t ← (target_image − image_pred)^2 − Δimage_pred
            Δimage_pred ← (target_image − image_pred)^2
        end for
        Δ(w_Qbot) ← (1/10) Σ_{t=1}^{10} ∇_{θ_Qbot} [G_t log p(ques_t, θ_Qbot) − Δimage_pred]
        Δ(w_Abot) ← (1/10) Σ_{t=1}^{10} G_t ∇_{θ_Abot} log p(ans_t, θ_Abot)
        w_Qbot ← w_Qbot + Δ(w_Qbot)                              ▷ REINFORCE and Image Loss update for Qbot
        w_Abot ← w_Abot + Δ(w_Abot)                              ▷ REINFORCE update for Abot
    end while
end procedure
```

Model	MRR	Mean Rank	R@1	R@5	R@10
Answer Prior (Das et al., 2016)	0.3735	26.50	23.55	48.52	53.23
MN-QIH-G (Das et al., 2016)	0.5259	17.06	42.29	62.85	68.88
HCIAE-G-DIS (Lu et al., 2017)	0.547	14.23	44.35	65.28	71.55
Frozen-Q-Multi (Das et al., 2017)	0.437	21.13	N/A	53.67	60.48
CoAtt-GAN (Wu et al., 2017)	0.5578	14.4	**46.10**	**65.69**	71.74
SL(Ours)	**0.610**	**5.323**	34.74	57.67	**72.68**
RL - 1Q,1A(Ours)	0.459	7.097	16.04	54.69	72.34
RL - 1Q,3A(Ours)	0.601	5.495	34.83	57.47	72.48
RL - 3Q,1A(Ours)	0.590	5.56	33.59	57.73	72.61

Table 1: Comparison of Metrics with Literature

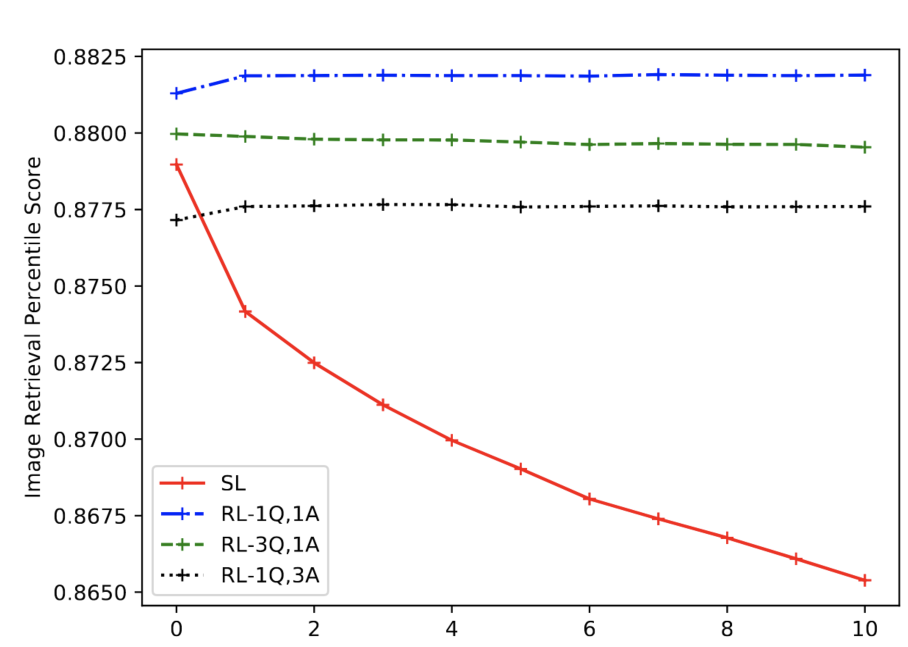


Figure 3: The percentile scores of the ground truth image compared to the entire test set of 40k images. The X-axis denotes the dialog round number (from 1 to 10), while the Y-axis denotes the image retrieval percentile score.

	Metric	N	Supervised	RL 1Q,1A	RL 1Q,3A	RL 3Q,1A
1	Q-Bot Relevance	8	2.5	2.75	**2**	2.75
2	Q-Bot Grammar	8	**2.25**	2.875	2.5	2.375
3	A-Bot Relevance	12	2.5	2.583	2.25	**1.67**
4	A-Bot Grammar	12	1.92	3.5	**1.83**	2.25
5	Overall Coherence	20	2.8	3.05	**2.3**	**1.85**

Table 2: Human Evaluation Results - Mean Rank (Lower is better). **N** refers to the number of human evaluators involved in the ranking.

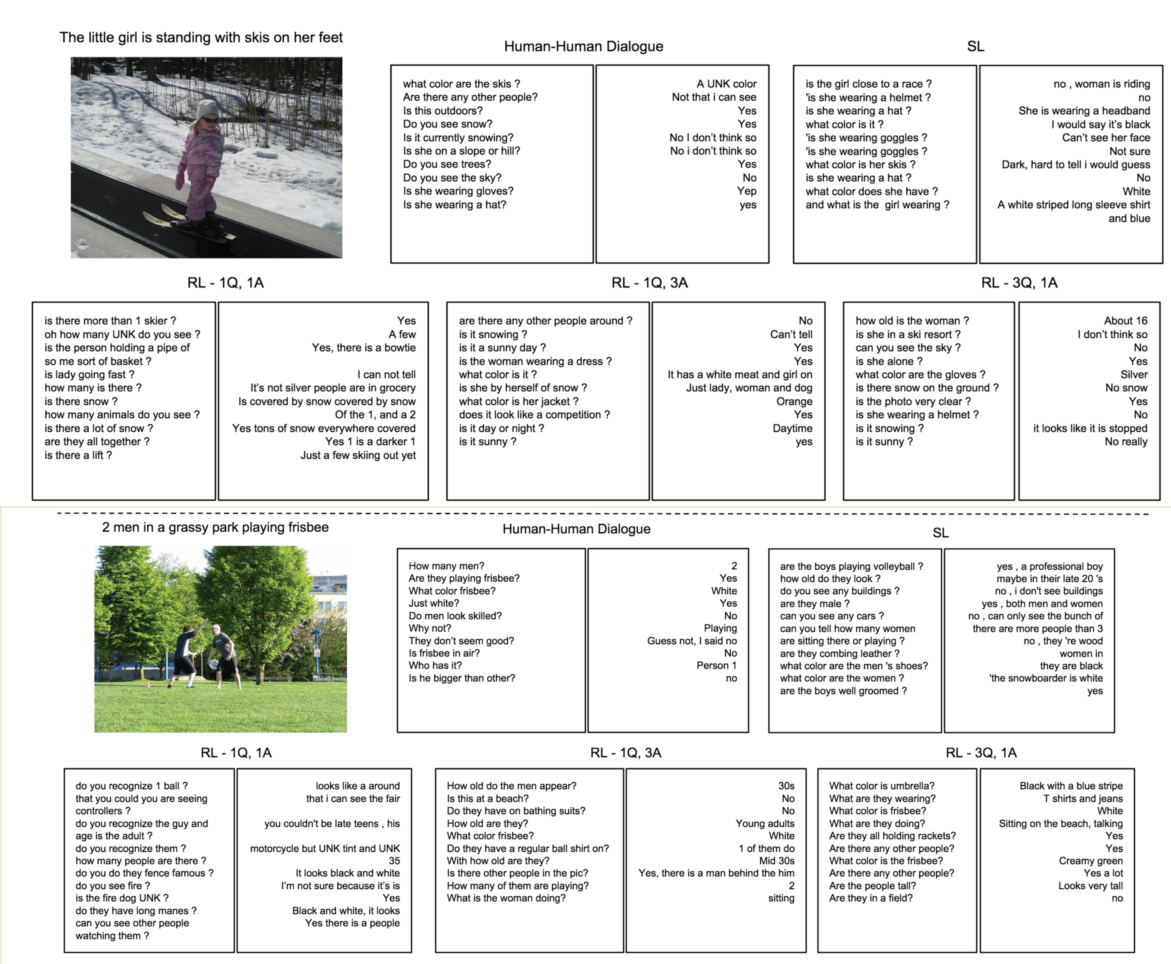


Figure 4: Two randomly selected images from the VisDial dataset followed by the ground truth (human) and generated dialog about that image for each of our 4 systems (SL, RL-1Q,1A, RL-1Q,3A, RL-3Q,1A). These images were also used in the human evaluation results shown in Table 2

Future Work

- We plan to explore several other multi bot architectural settings and perform a more thorough human evaluation for qualitative analysis of our dialog.
- We also plan on incorporating other language priors in our reinforcement learning setup to further improve the dialog quality.
- We will also experiment with using a discriminative answer decoder which uses information of the possible answer candidates to rank the generated answer with respect to all the candidate answers and use the ranking performance to train the answer decoder.

.

References

- Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, Jose M. F. Moura, Devi Parikh, and Dhruv Batra. 2016.Visual dialog.CoRR, abs/1611.08669
- Abhishek Das, Satwik Kottur, Jose M. F. Moura, StefanLee, and Dhruv Batra. 2017. Learning cooperativevisual dialog agents with deep reinforcement learn-ing.CoRR, abs/1703.06585.
- Jiasen Lu, Anitha Kannan, Jianwei Yang, Devi Parikh,and Dhruv Batra. 2017.Best of both worlds: Transferring knowledge from discriminative learn-ing to a generative visual dialog model. CoRR, abs/1706.01554
- Satwik Kottur, Jos´e M. F. Moura, Stefan Lee, andDhruv Batra. 2017.Natural language does notemerge 'naturally' in multi-agent dialog.CoRR,abs/1706.08502.
- Mike Lewis, Denis Yarats, Yann N Dauphin, DeviParikh, and Dhruv Batra. 2017. Deal or no deal?end-to-end learning for negotiation dialogues.arXivpreprint arXiv:1706.05125