Supplementary Material

In this section, we present supplementary experiemnts and results.

8. Details of Experiment Setting

For our Guesser model, the object states are initialized uniformly on the candidate objects in the images provided in the *GuessWhat?!* dataset. We use AdamW optimizer [18] with a learning rate of 1e-5. We usually observe convergence of the Guesser model after 3-4 epochs. For our Oracle model, we use AdamW optimizer with a learning rate of 1e-5. We usually observe convergence of the Oracle model after 3 epochs. For our Questioner model, we use AdamW optimizer with a learning rate of 1e-3. We also keep the state-estimator freezed during training.

9. Ablation Studies

9.1. Oracle Versus Guesser

In Table-9, we present examples where the following four combinations of Oracle and Guesser Model model are compared on the same test data: (1) Baseline Oracle + Baseline Guesser; (2) VilBERT Oracle + Baseline Guesser; (3) Baseline Oracle + VilBERT Guesser; (4) VilBERT Oracle + VilBERT Guesser. From the examples we can see that both Oracle and Guesser contribute to the final correct guessing. Neither a stronger Oracle nor a stronger Guesser alone can achieve a superior end-to-end performance, with the other two models fixed on baseline models. Only when both models are stronger and also understand each other, the final guessing is more likely to be correct.

9.2. A Weak Oracle

We also experimented with the setting of [baseline Oracle, VilBERT Guesser, VilBERT Questioner], and the end-to-end success rate dropped from 62.8% to 52.7%.

9.3. Using Advanced Language Encoder in Baseline Oracle Model

We experimented with the baseline Oracle model that uses the same object encoder (Faster-RCNN) [24] with a pre-trained BERT model [10] as language encoder. The final accuracy is 78.67%, which is lower than our proposed model (85.0%).

10. Reinforcement Learning Training

Although the present paper focuses on supervised learning on the *GuessWhat?!* task, we also experimented with reinforcement learning training of the agents. More specifically, we first train the three models, namely Guesser, Questioner and Oracle in supervised setting as described previously. Then we update the Questioner model with policy gradient using the **REINFORCE** algorithm [35]. After 3 epochs, the final end-game success rate on the same test set is 0.65. We haven't fully fine-tuned our Questioner model in this setting, due to the significant computation time to generate questions and update states with our VilBERT-based models and the vast number of sampling epochs required in reinforcement learning. We consider this as one of our major future work to propose more efficient sampling algorithm and and speed up the reinforcement learning training process.

Models	2-Gram	3-Gram	4-Gram
VDST	0.63	0.55	0.47
Our Questioner	**0.54**	**0.44**	**0.34**

Table 8: Self-Bleu for Language Diversity of Questioner Models (Lower is better).

11. Language Diversity of Generated Questions

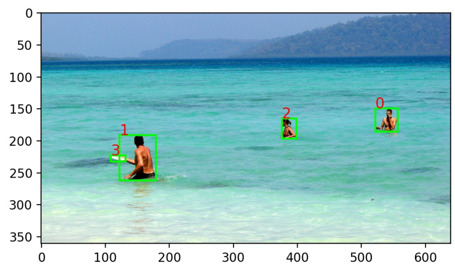
To evaluate language diversity of generated questions, we utilize **Self-Bleu** [43] to compare our VilBERT-Questioner and the state-of-the-art Questioner (VDST). The results are shown in Table 8. We can see that our Questioner model has higher language diversity compared to SOTA Questioner after supervised learning.

12. Successful Cases

Figure-6 - Figure-9 present four successful cases of proposed models. We show how object state beliefs are updated by our VilBERT Guesser model from turn to turn. You can see that the state estimator eliminates the unlikely objects according to the question and answer in each turn, and narrows down to the final target object. The accurate estimation of object states turn-by-turn can be largely attributed to the pretrained VilBERT encoder, which already shows state-of-the-art performance on referring expression comprehension tasks.

13. Failure Cases

Figure-10 - Figure-11 present two failure cases of proposed Guesser model. For Figure-10, we can see that the ground-truth dialog (between two humans) ends prematurely, in which both object-0 and object-4 (vehicles) are made of metal. The Guesser model has made a wrong guess based on the premature dialog to predict the target as object-4 while ground-truth target is object-0. In Figure-11, similar pattern can be observed. Both object-1 and object-2 satisfies the conditions: not a person; not clothing; not animal; has wheels and has handles.



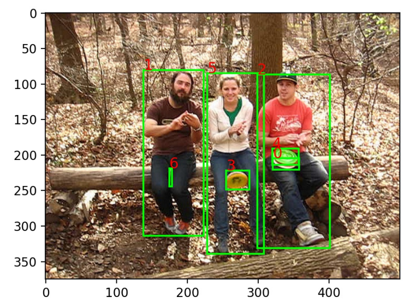
Ground-truth object: 2

Baseline Oracle + Baseline Guesser: 0
is it a person?—yes
is it the person on the left?—no
is it the person on the right?—yes

Vilbert Oracle + Baseline Guesser:1
is it a person?—yes
is it the person on the left?—no
is it the person on the right?—no

Baseline Oracle + Vilbert Guesser: 0
is it a person?—yes
is it the person on the left?—no
is it the person on the right?—yes

Vilbert Oracle + Vilbert Guesser:2
is it a person?—yes
is it the person on the left?—no
is it the person on the right?—no



Ground-truth object: 2

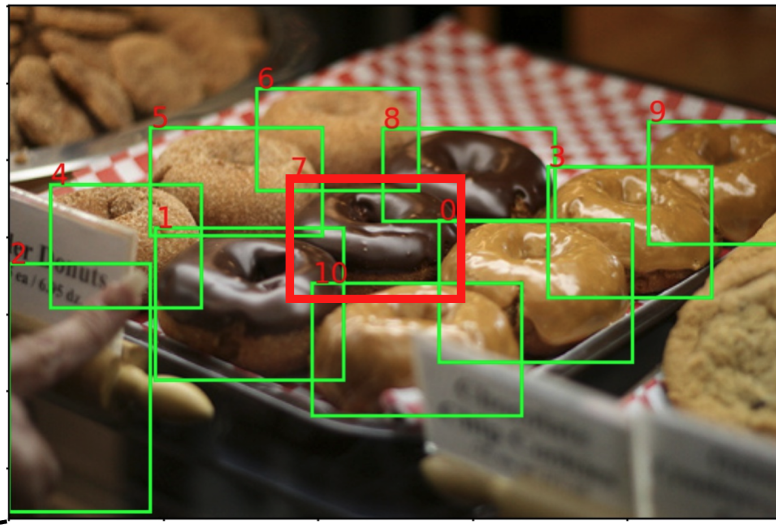
Baseline Oracle + Baseline Guesser: 1
is it a person?—yes
is it a female?—yes
is she wearing a hat?—yes

Vilbert Oracle + Baseline Guesser:1
is it a person?—yes
is it a female?—no
is it a man?—yes
is he wearing a hat?—yes

Baseline Oracle + Vilbert Guesser: 5
is it a person?—yes
is it a female?—yes
is she wearing a hat?—yes

Vilbert Oracle + Vilbert Guesser:2
is it a person?—yes
is it a female?—no
is it a man?—yes
is he wearing a hat?—yes

Table 9: Examples of Guesser Models Behavior Conditioned on Oracle Models



	obj-0	obj-1	obj-2	obj-3	obj-4	obj-5	obj-6	obj-7	obj-8	obj-9	obj-10
Initialization											
	9.1%	9.1%	9.1%	9.1%	9.1%	9.1%	9.1%	9.1%	9.1%	9.1%	9.1%
T-1 Question: is it a donut? Answer: Yes											
	9.1%	9.8%	0.9%	8.2%	6.6%	9.2%	8.3%	7.8%	9.9%	8.7%	10.0%
T-2 Question: does it have chocolate frosting? Answer: Yes											
	0.9%	30.9%	0.1%	0.8%	0.7%	0.9%	0.8%	11.1%	50.6%	0.9%	11.0%
T-3 Question: is it just one of them? Answer: Yes											
	1.3%	25.6%	0.3%	1.1%	0.8%	1.3%	1.2%	22.1%	42.3%	1.1%	1.6%
T-4 Question: is it in the middle? Answer: Yes											
	0.2%	8.7%	0.0%	0.1%	0.1%	0.2%	0.2%	51.3%	38.7%	0.1%	0.2%

Figure 6: Example-1 of Vilbert-Guesser State Estimator (Successful Case on Ground-Truth Dialogs)

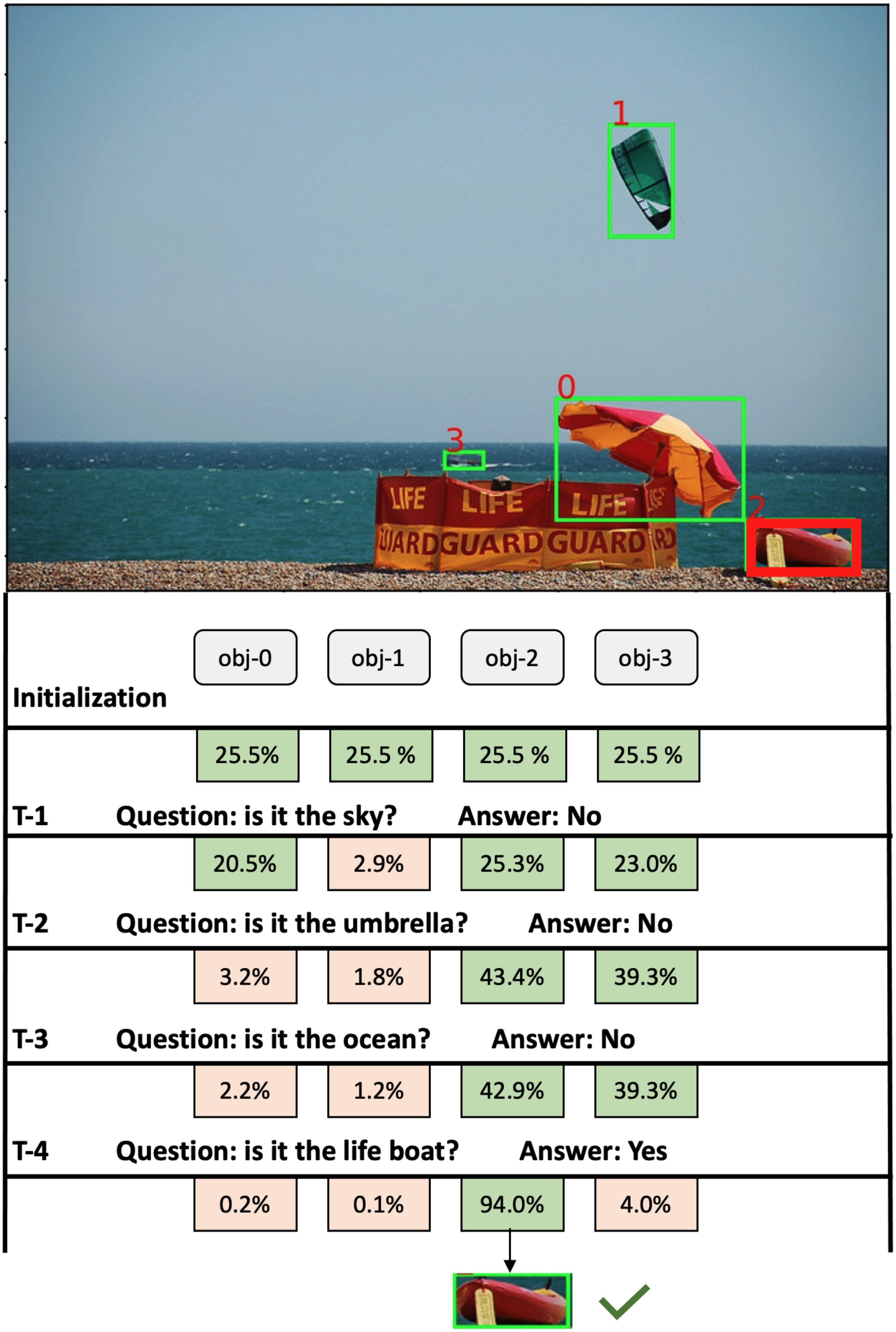


Figure 7: Example-2 of Vilbert-Guesser State Estimator (Successful Case on Ground-Truth Dialogs)

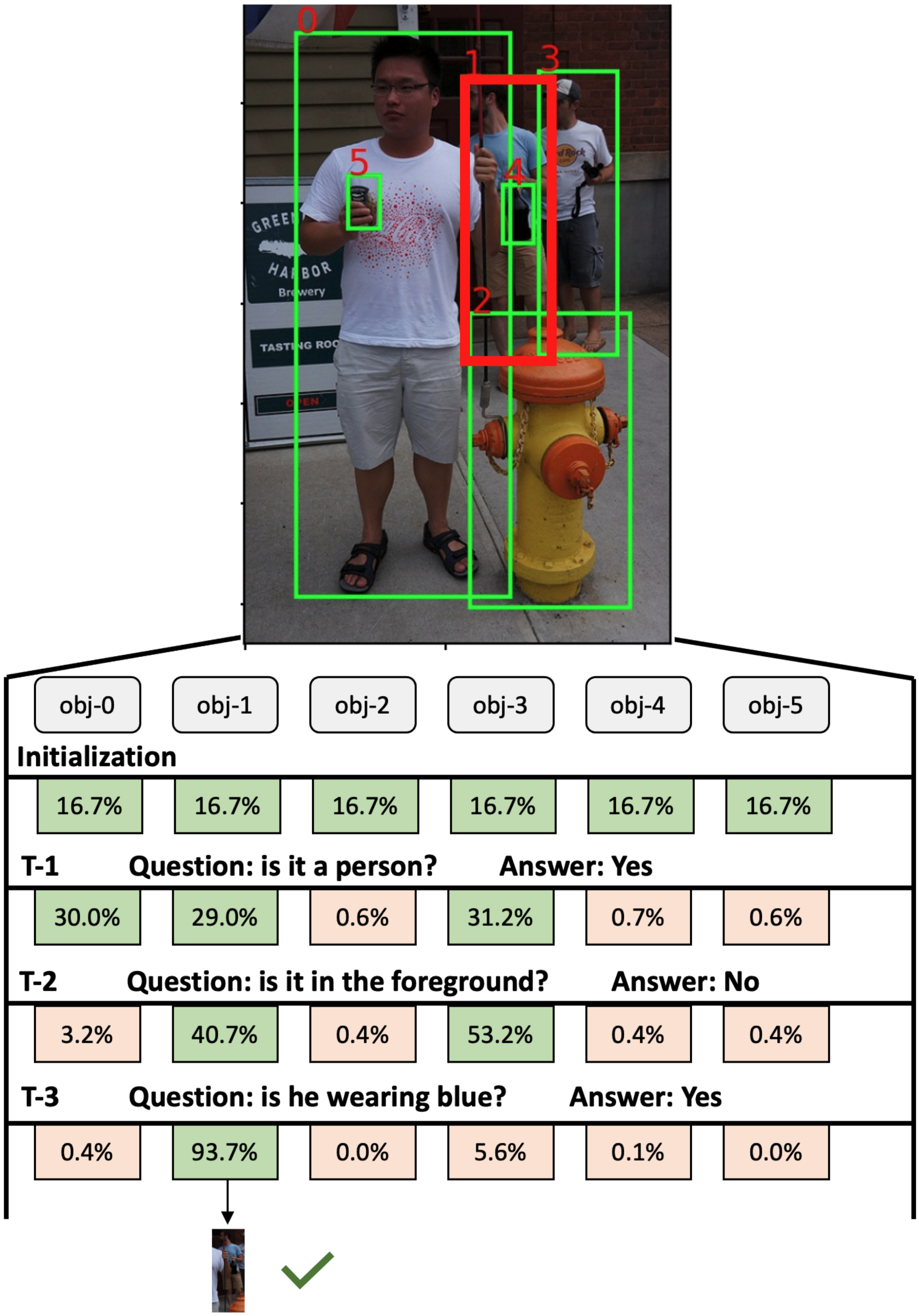


Figure 8: Example-3 of Vilbert-Guesser State Estimator (Successful Case on Ground-Truth Dialogs)

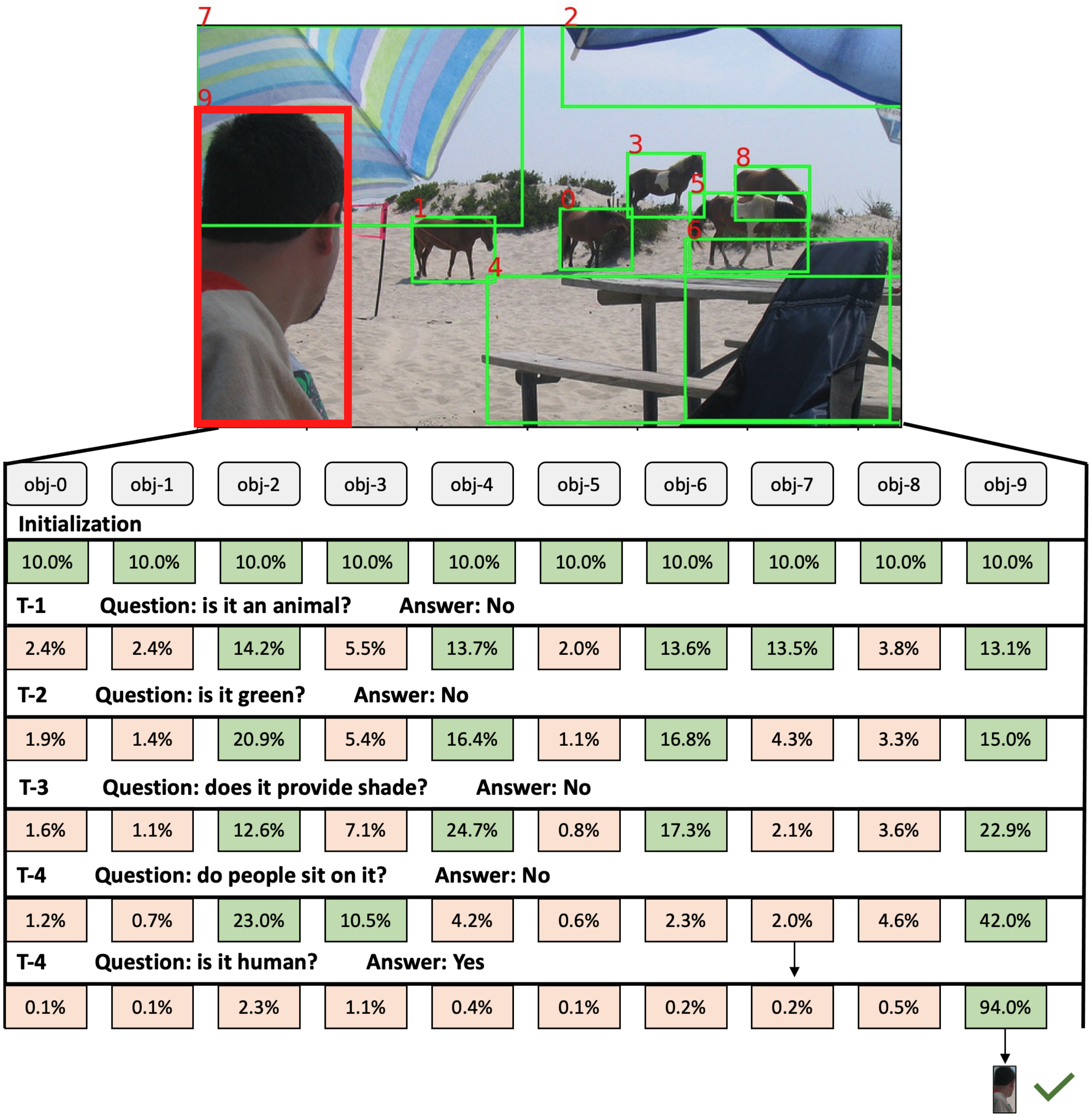


Figure 9: Example-4 of Vilbert-Guesser State Estimator (Successful Case on Ground-Truth Dialogs)

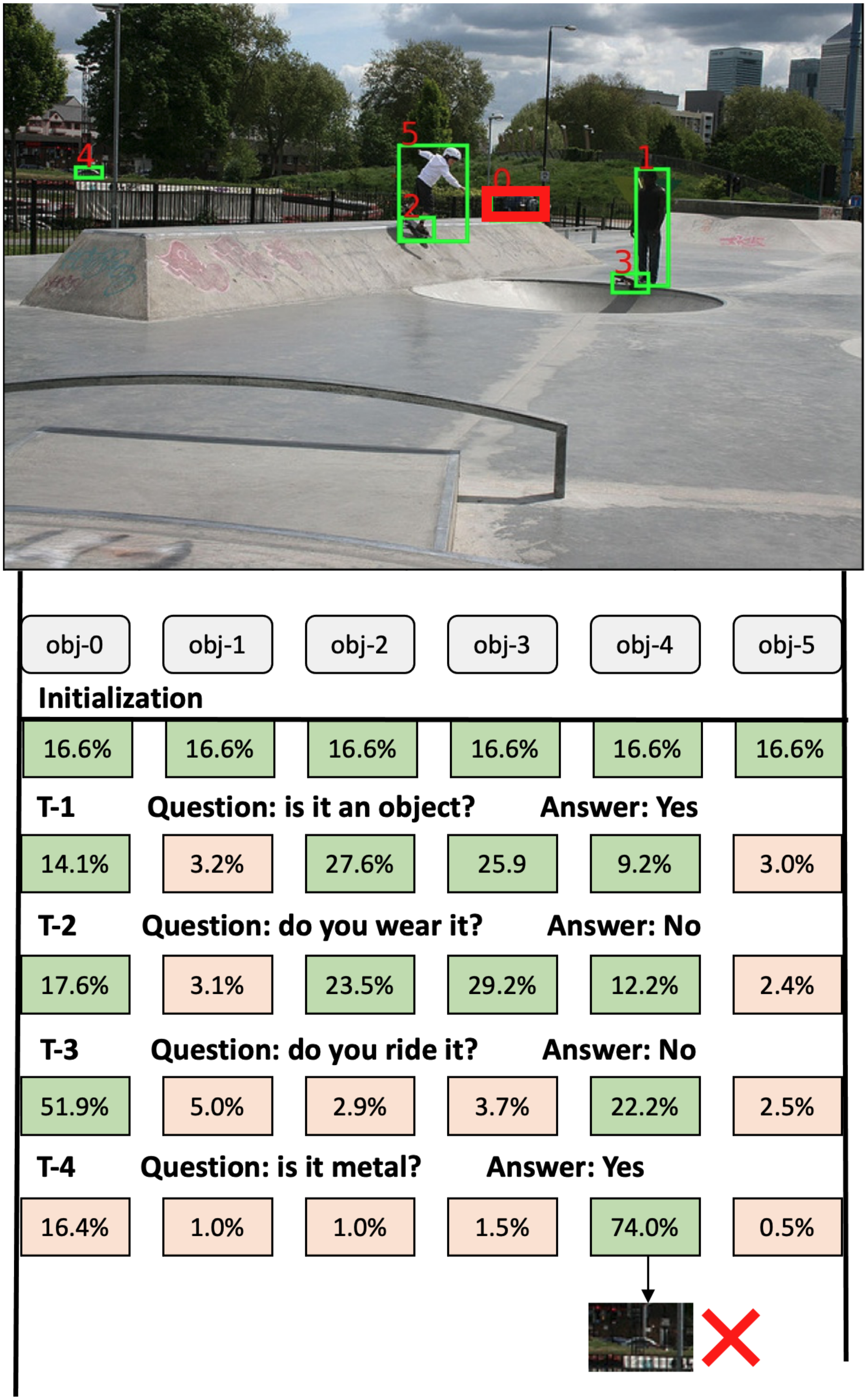


Figure 10: Example-1 of Vilbert-Guesser State Estimator (Failure Case on Ground-Truth Dialogs)

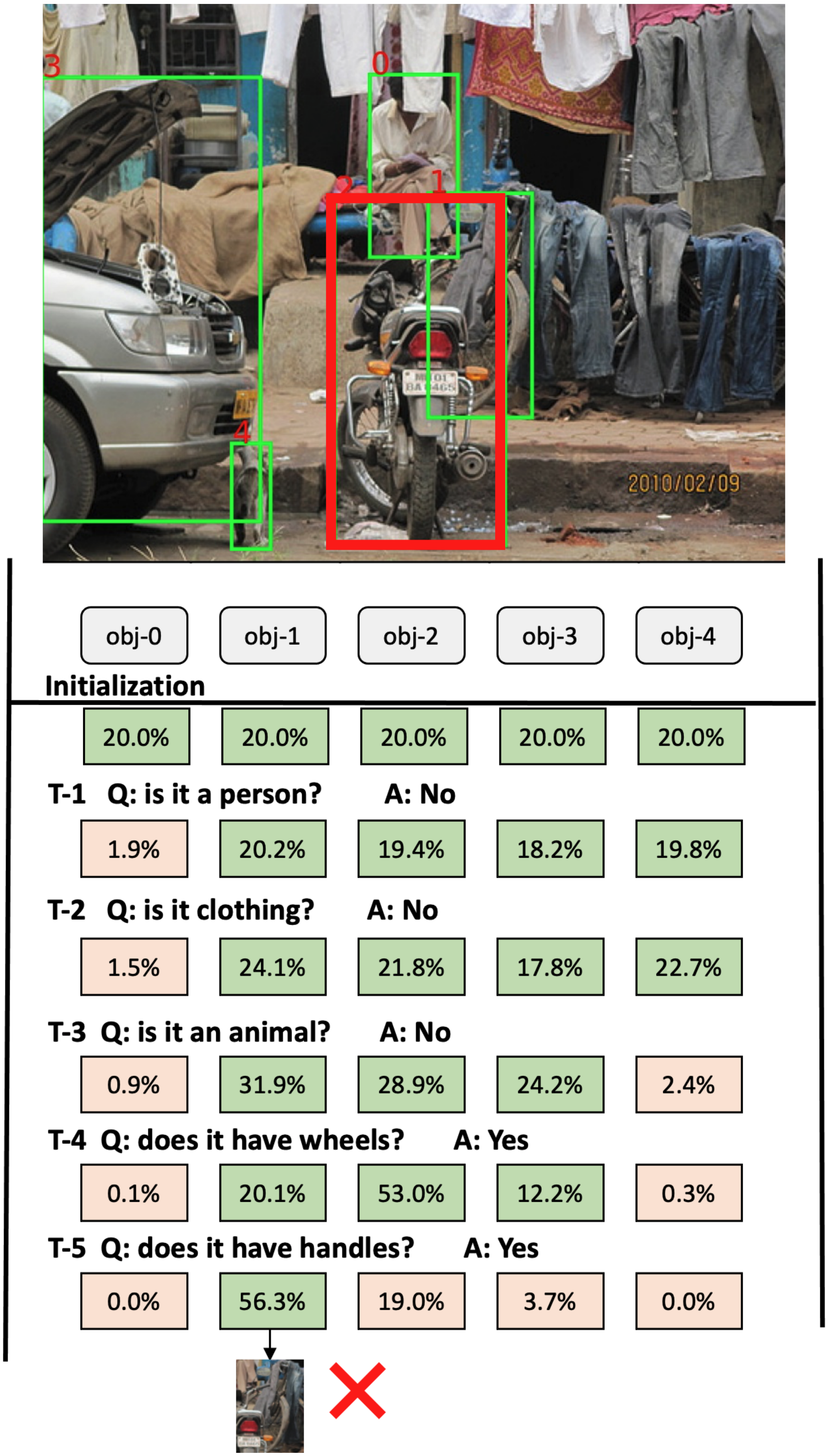


Figure 11: Example-2 of Vilbert-Guesser State Estimator (Failure Case on Ground-Truth Dialogs)