Bidirectional Beam Search: Forward-Backward Inference in Neural Sequence Models for Fill-in-the-Blank Image Captioning - Supplementary

Qing Sun
Virginia Tech
sunqing@vt.edu

Stefan Lee
Virginia Tech
steflee@vt.edu

Dhruv Batra
Georgia Tech
dbatra@gatech.edu

BiBS Convergence. As mentioned in the main paper, our proposed algorithm, BiBS, typically converges within in 1 to 2 rounds for the fill-in-the-blank image captioning task. Fig. 1 shows additional qualitative results that demonstrate how the highest ranked sentences change as the BiBS algorithm progress through these meta-iterations.

Unknown Length Blanks for Visual Madlibs. We extend BiBS on the Visual Madlib [1] fill-in-the-blank description generation tasks to the unknown length blanks setting. The basic setting is the same with main paper. We find that BiBS outperforms nearly all baselines on all metrics (narrowly being bested by GSN(ordered) at Blue-2 for type-7).

	type 7		type 12	
	Bleu-1	Bleu-2	Bleu-1	Bleu-2
URNN-f	0.317	0.155	0.285	0.174
URNN-b	0.334	0.184	0.309	0.186
URNN-f+b	0.334	0.181	0.302	0.184
BiRNN-f+b	0.343	0.195	0.291	0.190
GSN [2] (Ordered)	0.348	**0.203**	0.270	0.184
BiRNN-BiBS	**0.351**	0.197	**0.31**	**0.190**

Table 1: Unknown blank length setting on the Visual Madlibs task using BLEU-1 and BLEU-2. B=5 by default.

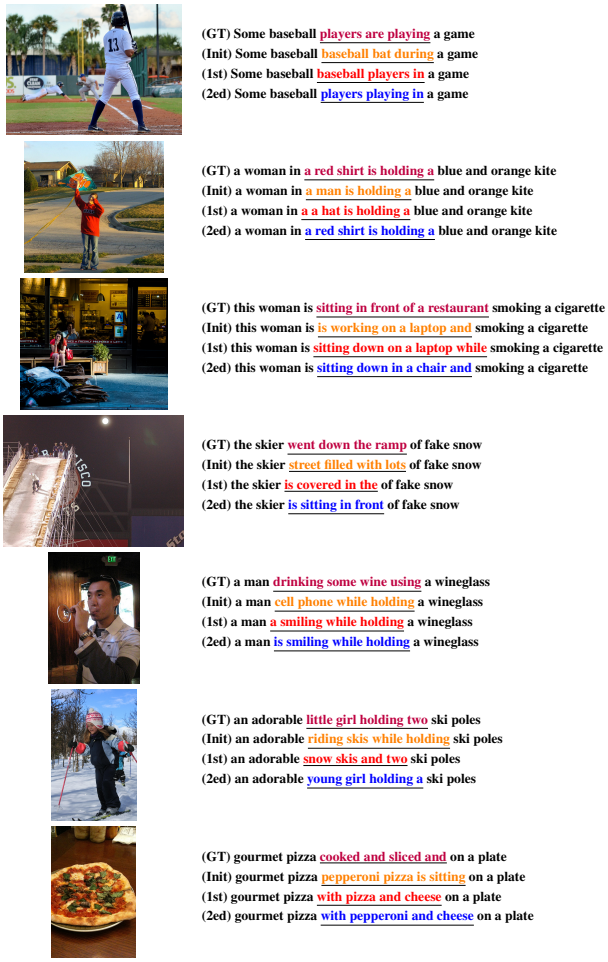

Figure 1: **Performance vs. Iteration.** Our model is initialized with right-to-left standard BS (Init) and updated alternatively from left-to-right (1st) and right-to-left (2nd).

References

[1] L. Yu, E. Park, A. C. Berg, and T. L. Berg, "Visual Madlibs: Fill in the blank Description Generation and Question Answering," *ICCV*, 2015. 1

[2] M. Berglund, T. Raiko, M. Honkala, L. Karkkainen, A. Vetek, and J. Karhunen, "Bidirectional recurrent neural networks as generative models," in *NIPS*, 2015. 1