ICCV
#2105

ICCV
#2105

ICCV 2017 Submission #2105. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Referring Expression Generation and Comprehension via Attributes
Supplementary Material

Anonymous ICCV submission

Paper ID 2105

1. Complete results evaluated by automated metrics

Table 1 shows the complete results of referring expression generation on RefCOCO, RefCOCO+ and RefCOCOg evaluated by automated metrics.

2. More comprehension results by automatic detection

In Figure 1 we show more comprehension results of "attr" based on the detected objects from Faster-RCNN [1]. The top three rows are the correct comprehension results and the bottom two rows are the incorrect ones. The green, blue and red boxes represent ground truth, correct and incorrect localizations respectively. The comprehension is successful if the overlap with the ground truth box is higher than 0.5. The failures are composed of two types: wrong comprehension by other objects and inaccurate localization.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detec-tion with region proposal networks. *In NIPS*, 2015. 1

[2] L. Yu, P. Poirson, S. Yang, A. C. Berg, and T. L. Berg. Modeling context in referring expressions. *In ECCV*, 2016. 2

Table 1. Referring expression generation results evaluated by automated metrics on RefCOCO, RefCOCO+ and RefCOCOg.

RefCOCO

	TestA				TestB			
	BLEU1	BLEU2	ROUGE	METEOR	BLEU1	BLEU2	ROUGE	METEOR
baseline [2]	0.477	0.290	0.413	0.173	0.553	0.343	0.499	0.228
MMI [2]	0.478	0.295	0.418	0.175	0.547	0.341	0.497	0.228
visdif+MMI [2]	0.494	0.307	0.441	0.185	0.578	0.375	0.531	0.247
visdif+tie [2]	0.510	0.318	0.446	0.189	0.593	0.386	0.533	0.249
attr	0.553	0.366	0.472	0.208	0.581	0.379	0.532	0.247
attr+visdif	**0.562**	**0.373**	**0.494**	**0.222**	**0.606**	**0.392**	**0.546**	**0.258**

RefCOCO+

	TestA				TestB			
	BLEU1	BLEU2	ROUGE	METEOR	BLEU1	BLEU2	ROUGE	METEOR
baseline [2]	0.391	0.218	0.356	0.140	0.331	0.174	0.322	0.135
MMI [2]	0.370	0.203	0.346	0.136	0.324	0.167	0.320	0.133
visdif+MMI [2]	0.386	0.221	0.360	0.142	0.327	0.172	0.325	0.135
visdif+tie [2]	0.409	0.232	0.372	0.150	0.340	0.178	0.328	0.143
attr	0.399	0.229	0.362	0.150	0.342	0.181	0.345	0.149
attr+visdif	**0.415**	**0.239**	**0.374**	**0.155**	**0.350**	**0.189**	**0.355**	**0.155**

RefCOCOg

	Val			
	BLEU1	BLEU2	ROUGE	METEOR
baseline [2]	0.437	0.273	0.363	0.149
MMI [2]	0.428	0.263	0.354	0.144
visdif [2]	0.442	0.277	0.370	0.151
attr	**0.460**	**0.295**	**0.389**	**0.163**
attr+visdif	0.454	0.291	0.378	0.160

ICCV
#2105

ICCV
#2105

ICCV 2017 Submission #2105. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

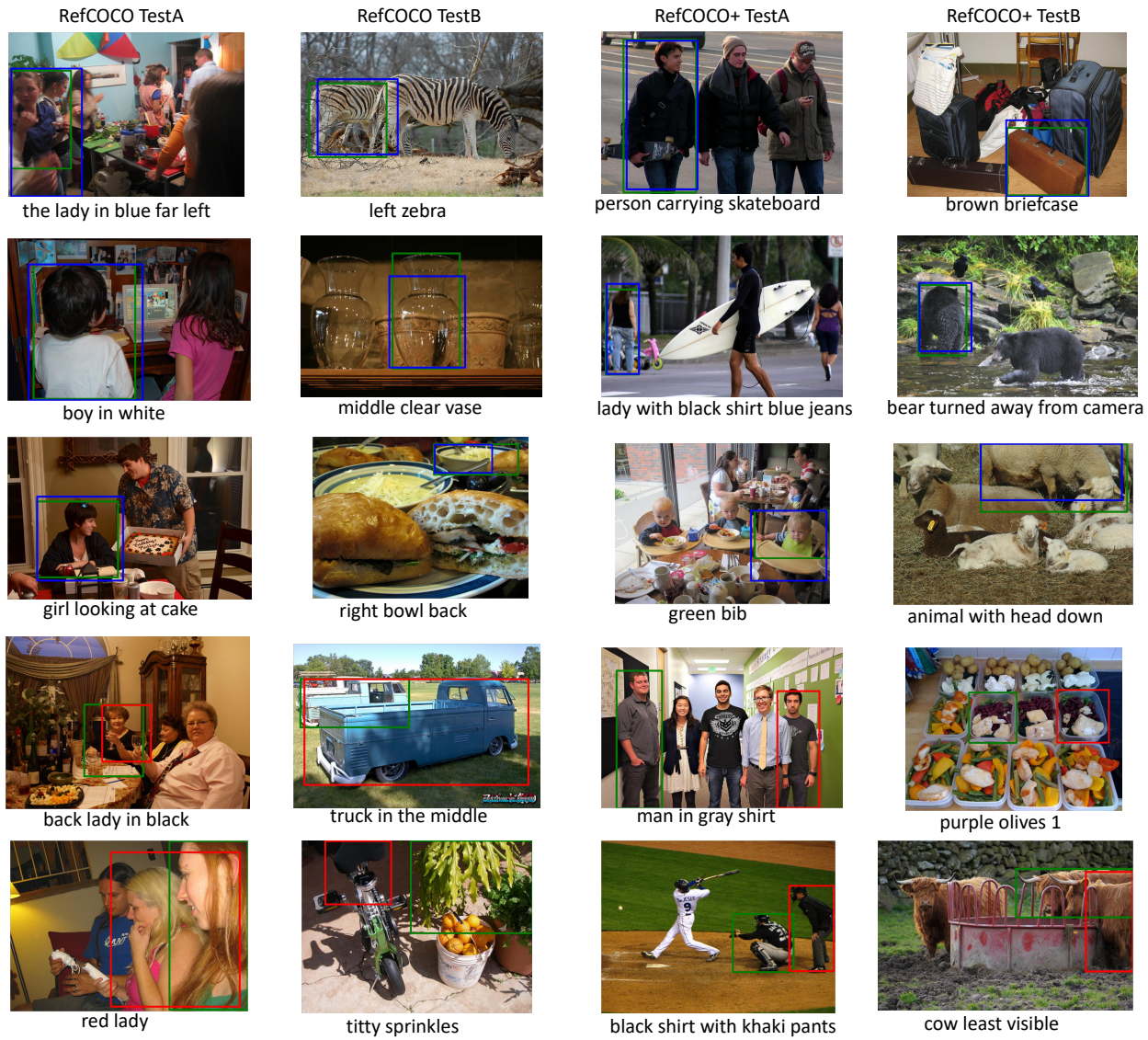| RefCOCO TestA	RefCOCO TestB	RefCOCO+ TestA	RefCOCO+ TestB
the lady in blue far left	left zebra	person carrying skateboard	brown briefcase
boy in white	middle clear vase	lady with black shirt blue jeans	bear turned away from camera
girl looking at cake	right bowl back	green bib	animal with head down
back lady in black	truck in the middle	man in gray shirt	purple olives 1
red lady	titty sprinkles	black shirt with khaki pants	cow least visible

Figure 1. Some comprehension results of "attr" based on the detected objects from Faster-RCNN. The green, blue and red boxes represent ground truth, correct and incorrect localizations respectively.