Question Answering about Images using Visual Semantic Embeddings

Table 1. COCO-QA Question Type Break-Down

Category	Train	%	Test	%
Object	58352	69.96%	27497	71.13%
Number	6551	7.85%	2575	6.66%
Color	13337	15.99%	6047	15.64%
Location	5164	6.19%	2541	6.57%
Total	83404	100.00%	38660	100.00%

Table 2. DAQUAR Results

	Acc.	WUPS 0.9
2-Cnn-Lstm	**0.3578**	**0.3602**
Cnn-Lstm	0.3441	0.3464
Blind-Lstm	0.3273	0.3294
Blind-Bow	0.3267	0.3289
Guess	0.1824	0.1671
Multi-World	0.1273	0.1810
Human	0.6027	0.6104

Table 3. COCO-QA Results

	Acc.	WUPS 0.9
2-Cnn-Lstm	**0.5161**	**0.5244**
Cnn-Lstm	0.5073	0.5153
Img-Bow	0.4490	0.4593
Blind-Lstm	0.3516	0.3592
Blind-Bow	0.3262	0.3337
Guess	0.0665	0.0851

Table 4. COCO-QA Accuracy Per Category Break-Down

	Object	Number	Color	Location
2-Cnn-Lstm	**0.5386**	0.4534	**0.4786**	0.4258
Cnn-Lstm	0.5321	**0.4678**	0.4439	**0.4294**
Cnn-Bow	0.4718	0.3773	0.4130	0.3609
Blind-Lstm	0.3459	0.4383	0.3334	0.3680
Blind-Bow	0.3201	0.3339	0.3434	0.3436
Guess	0.0211	0.3584	0.1387	0.0893