

Attention Correctness in Neural Image Captioning (Supplementary Material)

Chenxi Liu¹ Junhua Mao² Fei Sha^{2,3} Alan Yuille^{1,2}
 Johns Hopkins University¹
 University of California, Los Angeles²
 University of Southern California³

In this supplementary material, we first present the spearsman correlation to show the legitimacy of our attention correctness metric, then provide more visualization in the generated caption experiment.

The Spearsman Correlation

Table 1: The spearsman correlation of potential correctness metrics on the Flickr30k test set with implicit attention and ground truth caption. Attention correctness correlates well with the other metrics.

	AC	L1	L2	KL
AC	1.0000	0.9890	0.9624	0.9838
L1		1.0000	0.9822	0.9958
L2			1.0000	0.9827
KL				1.0000

Attention Map Visualization

We visualize the attention maps of both the implicit attention model and our supervised attention model on the Flickr30k test set. As mentioned in the paper, 909 noun phrases are aligned for the implicit model and 901 for the supervised model. 635 of these alignments are common for both, and 595 of them have corresponding bounding boxes. Here we present a subset due to space.

For every figure, the original image is on the left, the implicit attention result is in the middle, and the supervised attention result is on the right. The red box marks the ground truth attention region, as annotated in the Flickr30k Entities dataset. The attention correctness score for this phrase is in the parenthesis.



Figure 1: *people* (0.242). *people* (0.477).

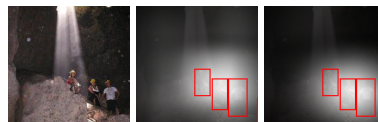


Figure 2: *people* (0.257). *people* (0.457).



Figure 3: *people* (0.637). *people* (0.822).

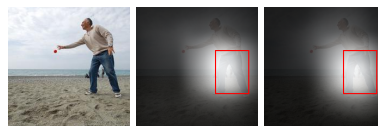


Figure 4: *jeans* (0.386). *jeans* (0.487).



Figure 5: *jeans* (0.320). *jeans* (0.320).



Figure 6: *people* (0.279). *people* (0.453).



Figure 7: *people* (0.564). *people* (0.602).



Figure 8: *people* (0.777). *people* (0.917).



Figure 9: *people* (0.486). *people* (0.656).



Figure 10: *people* (0.514). *people* (0.620).



Figure 11: *food* (0.060). *food* (0.265).



Figure 12: *jeans* (0.469). *jeans* (0.548).



Figure 13: *children* (0.281). *children* (0.328).



Figure 14: *men* (0.925). *men* (0.962).

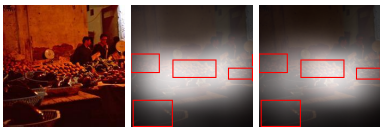


Figure 15: *vegetables* (0.292). *vegetables* (0.349).

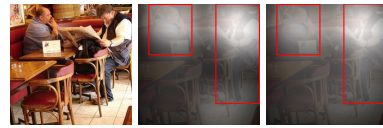


Figure 16: *people* (0.702). *people* (0.664).

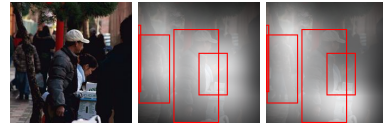


Figure 17: *people* (0.569). *people* (0.546).



Figure 18: *people* (0.660). *people* (0.616).

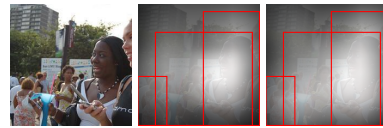


Figure 19: *people* (0.829). *people* (0.875).

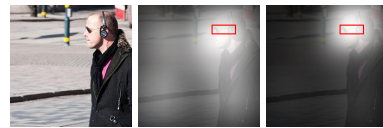


Figure 20: *sunglasses* (0.049). *sunglasses* (0.099).

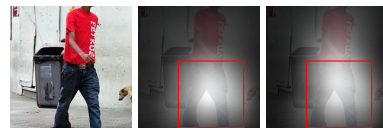


Figure 21: *jeans* (0.680). *jeans* (0.658).



Figure 22: *people* (0.250). *people* (0.404).



Figure 23: *people* (0.176). *people* (0.220).



Figure 24: *people* (0.329). *people* (0.441).



Figure 25: *bikes* (0.449). *bikes* (0.495).

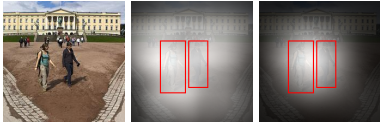


Figure 26: *people* (0.212). *people* (0.488).



Figure 27: *stage* (0.211). *stage* (0.301).



Figure 28: *people* (0.577). *people* (0.467).



Figure 29: *people* (0.608). *people* (0.627).



Figure 30: *food* (0.060). *food* (0.072).



Figure 31: *stage* (0.340). *stage* (0.480).

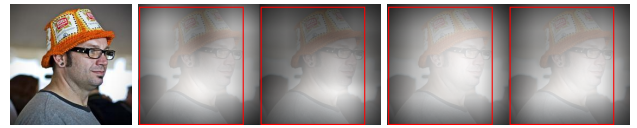


Figure 32: *A man* (0.865). *A man* (0.863).



Figure 33: *the snow* (0.390). *the snow* (0.461).



Figure 34: *A man* (0.072). *A man* (0.238).

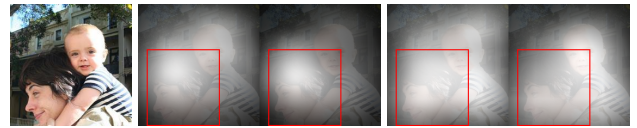


Figure 35: *A woman* (0.611). *a woman* (0.458).



Figure 36: *A man* (0.316). *A man* (0.574).



Figure 37: *A man* (0.406). *A man* (0.691).

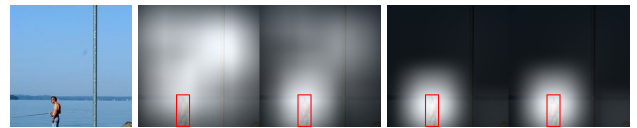


Figure 38: *A man* (0.104). *A man* (0.302).



Figure 39: *A man* (0.310). *A man* (0.573).



Figure 40: A man (0.570). A man (0.876).

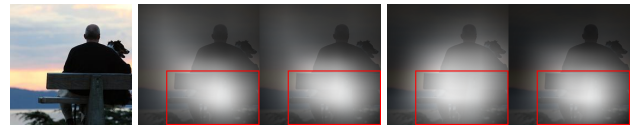


Figure 48: a bench (0.620). a bench (0.749).

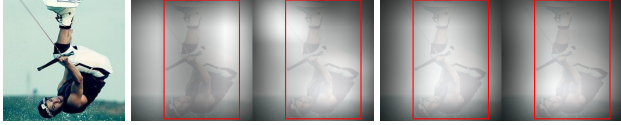


Figure 41: A man (0.666). A man (0.763).



Figure 49: a playground (0.507). a playground (0.485).



Figure 42: A man (0.239). A man (0.490).



Figure 50: a bench (0.571). a bench (0.479).

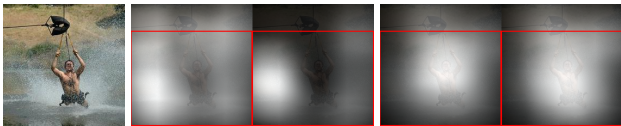


Figure 43: the water (0.894). the water (0.828).

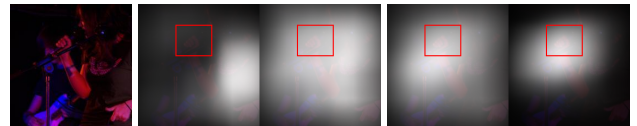


Figure 51: a microphone (0.100). a microphone (0.285).

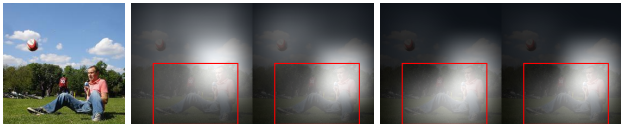


Figure 44: A man (0.616). A man (0.805).



Figure 52: Three men (0.628). Three men (0.678).



Figure 45: blue jeans (0.400). blue jeans (0.471).

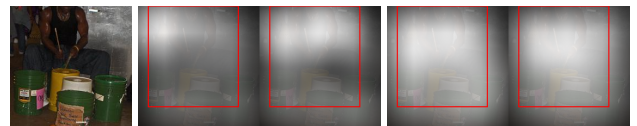


Figure 53: A man (0.643). A man (0.677).

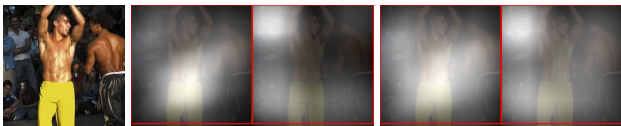


Figure 46: a crowd (0.990). a crowd (0.985).

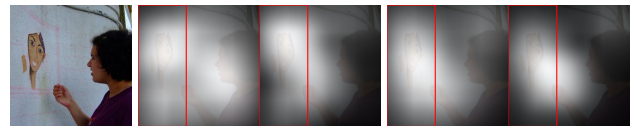


Figure 54: a wall (0.578). a wall (0.548).

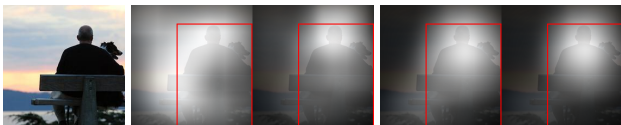


Figure 47: A man (0.701). A man (0.809).

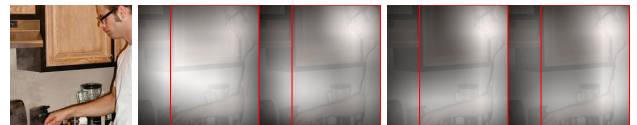


Figure 55: A man (0.799). A man (0.833).

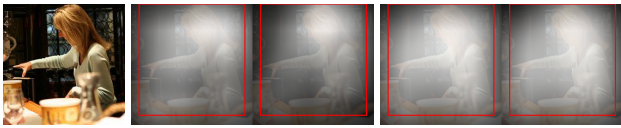


Figure 56: *A woman* (0.905). *A woman* (0.861).

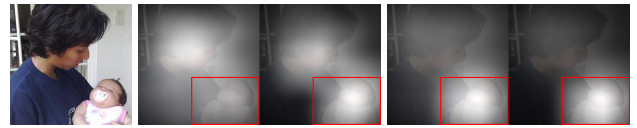


Figure 64: *a baby* (0.400). *a baby* (0.562).



Figure 57: *A man* (0.605). *A man* (0.635).



Figure 65: *a woman* (0.461). *a woman* (0.555).

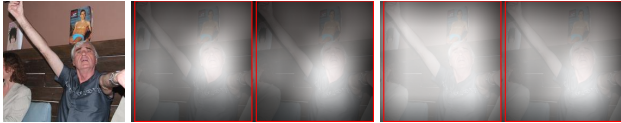


Figure 58: *A man* (0.977). *A man* (0.968).



Figure 66: *A man* (0.311). *A man* (0.570).



Figure 59: *a bulldozer* (0.756). *a bulldozer* (0.874).



Figure 67: *A man* (0.031). *A man* (0.054).



Figure 60: *A man* (0.261). *A man* (0.635).

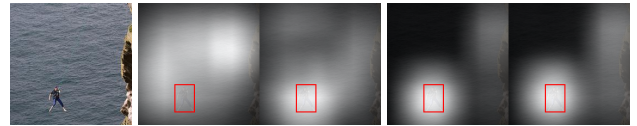


Figure 68: *A man* (0.068). *A person* (0.199).

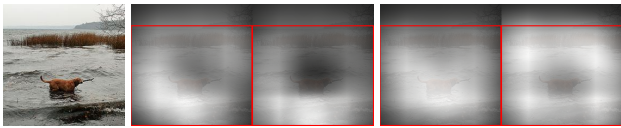


Figure 61: *the water* (0.933). *the water* (0.964).



Figure 69: *a bench* (0.646). *a bench* (0.748).



Figure 62: *A man* (0.460). *A man* (0.575).

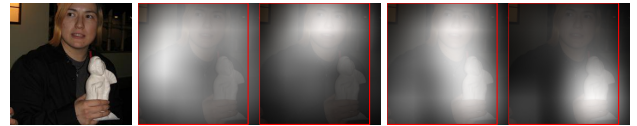


Figure 70: *A woman* (0.950). *a woman* (0.972).



Figure 63: *Two dogs* (0.224). *Two dogs* (0.438).



Figure 71: *the water* (0.995). *the water* (0.997).

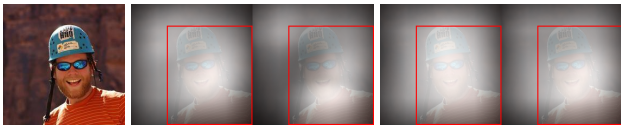


Figure 72: A man (0.628). A man (0.660).

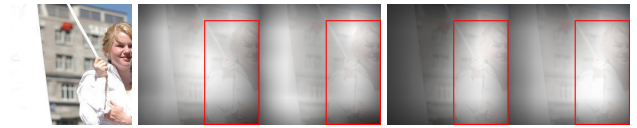


Figure 80: A woman (0.387). A woman (0.634).



Figure 73: A woman (0.812). A woman (0.869).

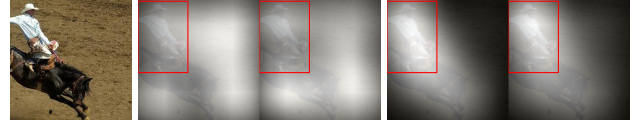


Figure 81: A man (0.204). A man (0.499).



Figure 74: a camera (0.101). a camera (0.216).



Figure 82: Construction workers (0.065). Construction workers (0.085).

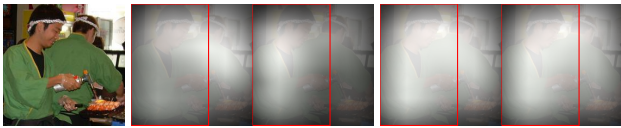


Figure 75: A man (0.725). A man (0.687).



Figure 83: a woman (0.686). A woman (0.731).



Figure 76: A dog (0.797). A dog (0.743).

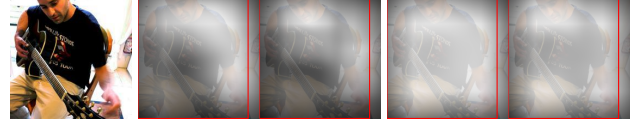


Figure 84: A man (0.919). A man (0.920).



Figure 77: a bench (0.508). a bench (0.582).



Figure 85: A man (0.199). A man (0.585).

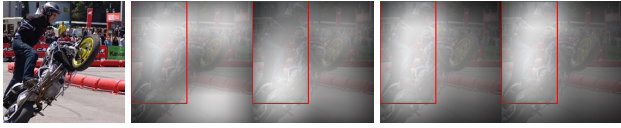


Figure 78: A man (0.463). A man (0.589).



Figure 86: A dog (0.757). A dog (0.739).

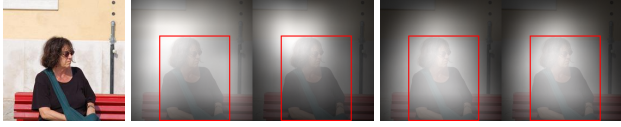


Figure 79: A woman (0.480). A woman (0.690).



Figure 87: a hurdle (0.802). a hurdle (0.826).

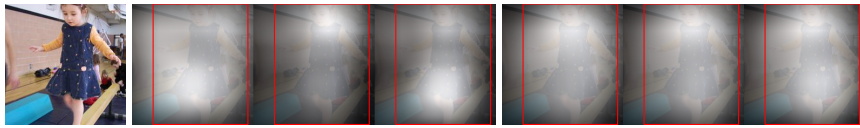


Figure 88: *A little girl* (0.847). *A little girl* (0.868).



Figure 89: *A young boy* (0.231). *A young boy* (0.434).

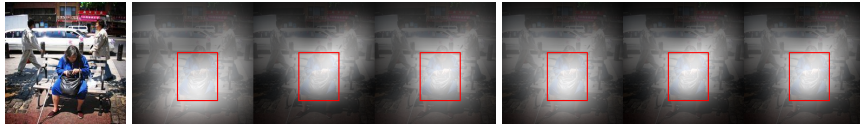


Figure 90: *a blue shirt* (0.411). *a blue shirt* (0.450).



Figure 91: *a blue shirt* (0.579). *a blue shirt* (0.706).



Figure 92: *A brown dog* (0.985). *A brown dog* (0.955).



Figure 93: *a black shirt* (0.308). *a black shirt* (0.401).

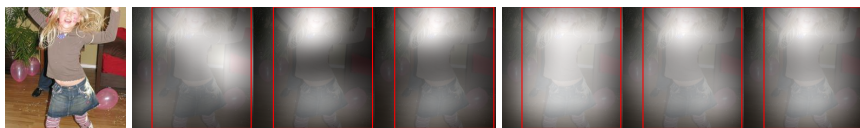


Figure 94: *A little girl* (0.932). *A little girl* (0.925).



Figure 95: *A brown dog* (0.338). *A dog* (0.403).



Figure 96: *a blue shirt* (0.559). *a blue shirt* (0.661).



Figure 97: *A black dog* (0.708). *A black dog* (0.760).

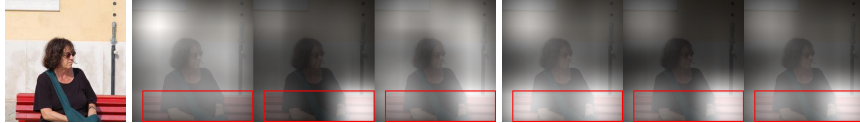


Figure 98: *a red bench* (0.500). *a red bench* (0.601).

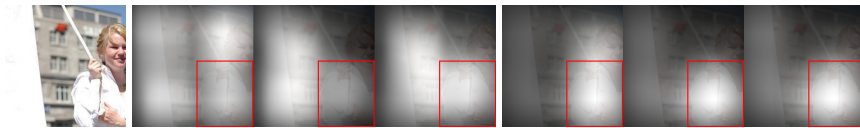


Figure 99: *a white shirt* (0.352). *a white shirt* (0.496).



Figure 100: *A black dog* (0.613). *A black dog* (0.566).



Figure 101: *a pink dress* (0.416). *a pink dress* (0.638).

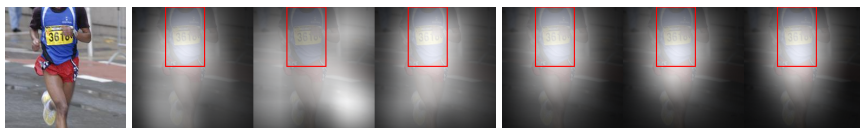


Figure 102: *a blue shirt* (0.314). *a blue shirt* (0.471).



Figure 103: *a white shirt* (0.842). *a white shirt* (0.874).



Figure 104: A group of people (0.799). A group of people (0.766).



Figure 105: A group of people (0.735). A group of people (0.952).



Figure 106: A group of children (0.924). A group of children (0.922).

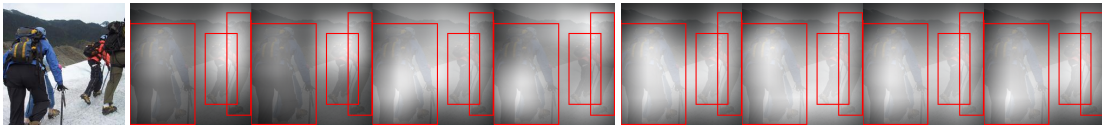


Figure 107: A group of people (0.842). A group of people (0.882).



Figure 108: A group of people (0.723). A crowd (0.828).



Figure 109: A group of people (0.717). A group of people (0.840).



Figure 110: A group of people (0.356). A group of people (0.396).



Figure 111: A group of people (0.756). A group of people (0.760).



Figure 112: *A group of people* (0.259). *A group of people* (0.444).



Figure 113: *A group of people* (0.634). *A group of people* (0.792).



Figure 114: *A group of people* (0.568). *A group of people* (0.634).

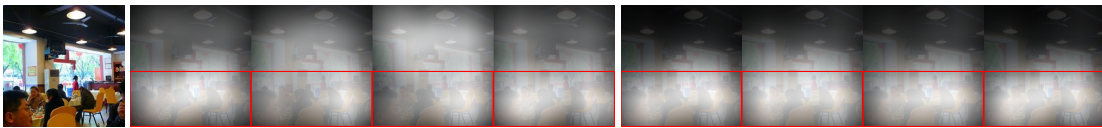


Figure 115: *A group of people* (0.665). *A group of people* (0.842).



Figure 116: *A group of people* (0.826). *A group of people* (0.840).



Figure 117: *A group of people* (0.869). *A group of people* (0.905).



Figure 118: *A group of people* (0.207). *A group of people* (0.265).



Figure 119: *A group of people* (0.323). *A group of people* (0.442).