Answer each question with a few sentences to a paragraph. Please e-mail answers in pdf format to mm@pdx.edu.

Questions on O. Vinyals et al., “Show and tell: A neural image caption generator”

1. Explain what the terms $W_c$ and $S_i$ are in Equation 11.

2. Explain briefly how Beam Search is used in generating a sentence from the trained model.

3. Briefly describe one specific thing you did not understand in this paper (citing section or page number if possible). I will collect these questions and try to answer them in class.

Questions on A. Karpathy & L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions

1. In the approach described in this paper, an image is represented as a set of $h$-dimensional vectors, $\{v_i | i = 1…20\}$. Explain where the number 20 comes from, and what each of the 20 $h$-dimensional vectors represent.

2. Explain what $S_{kl}$, defined in equation 8, represents, and explain the various terms in equation 8: $g_l$, $g_k$, $v_i$, and $s_l$.

3. In Section 4.1, the authors state that “An appealing feature of our model is that it leans to modulate the magnitude of the region and word embeddings.” Explain briefly what they mean by this sentence, and why this is a useful feature for their model.

4. Again, briefly describe one specific thing you did not understand in this paper (citing section or page number if possible). I will collect these questions and try to answer them in class.