Project Work - Image description generation

Deep Learning Winter School, University of Hull
January 23-24 2018
Dr Nina Dethlefs

The GRE3D7 dataset [1] is a set of images displaying objects in a variety of spatial arrangements alongside natural language description of a target object. An example is shown in Figure 1 and might come with an example "the small blue sphere in front of the large green cube". The dataset has 4,480 examples split into a training and a test set in a ration of 80%-20%.

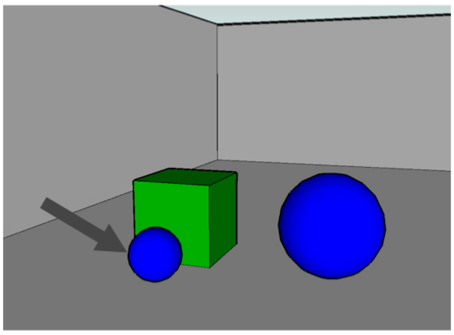

Figure 1: Example of the GRE3D7 scenes.

Exercise 2: Generating image descriptions. The idea is to develop a system that can generate descriptions for unseen scenes automatically based on a set of patterns that it learns during training. We will initially use the semantic patterns as input to our algorithm and then move on to using the images directly.

1. Run the given script *seq2seq_words.py* as it is. Observe how well the model learns its task. If there are any errors or inconsistencies, what are they and why might they occur?
2. Use the operations in *image_operations.py* to obtain a pixel matrix from an image and pass the pixel matrix into the neural net instead of the semantic pattern (e.g. "tg_size tg_col tg_type"). Do to this, you will need to load the relevant image (*GRE3D7-descriptions.csv* will tell you which), represent it in the correct way and change the input expectations of the neural network (e.g. the number of input nodes and the shape of the input, or its encoding).
3. Once this runs, observe the effects on the results. Are they better now? Why, why not?
4. Try to change some of the input images, e.g. by flicking a certain number of them, or changing their colours – anything you like. What impact does this have on the results - and why?
5. Try to keep track of the model's loss over time and generate a plot using *matplotlib*.

References

[1] J. Viethen and R. Dale, "Gre3d7: A corpus of distinguishing descriptions for objects in visual scenes," in *Proceedings of the UCNLG+Eval: Language Generation and Evaluation Workshop*, 2011, pp. 12–22.