**City Poetry**

Eital Schattner-Elmaleh, Charles Windolf

**Objectives**

This project will combine image captioning with generative language modeling. The goal is to produce two linked neural networks. One will take in images of street scenes in Providence, randomly obtained using the Google Maps API, and produce sensible captions for that image. The other network will take in those captions and generate poems using the captions as a seed.

**Current Methods and Datasets**

*Image Captioning:*

For the image captioning portion, we will use state of the art captioning models and datasets. What follows are some sources that we’re looking at, although we plan to either use the first (im2txt) either as-is or by re-implementing the same architecture with the minimal feature set for our use case, so that we can understand it better. In either case, we will retrain it on the data that’s relevant for our purposes, to avoid the intrusion of strange labels that won’t make sense in a street-scene context.

Models:

* Google’s open-sourced image captioning architecture, **im2txt** ([*source*](https://github.com/tensorflow/models/tree/master/im2txt)) ([*paper*](https://arxiv.org/abs/1609.06647))
* Andrej Karpathy’s **neuraltalk2** ([*source*](https://github.com/karpathy/neuraltalk2)) (neuraltalk1 [*source*](https://github.com/karpathy/neuraltalk)) ([*paper*](http://cs.stanford.edu/people/karpathy/deepimagesent/))
  + This one inspires **im2txt**, which improves on its results and training times.

Data Sources:

* Karpathy’s Flickr8K, Flickr30K, COCO datasets (region annotations and VGG weights), available in paper link above.
* MIT CSAIL LabelMe street scene region annotations ([*link*](http://labelme.csail.mit.edu/Release3.0/browserTools/php/publications.php))
* MIT CSAIL low-dimensional ‘spatial envelope’ scene representations ([*link*](http://people.csail.mit.edu/torralba/code/spatialenvelope/)). Has labels like ‘tall buildings’, ‘inside city’, ‘open country’, ‘highway’, which will be nice for poetry generation.
  + **Note:** we’re considering running this model in parallel to **im2txt** to generate part of the seed, instead of incorporating its labels into our **im2txt** training set.

*Generative Language Model:*

We plan to re-implement Andrej Karpathy’s **char-rnn** ([*expository explanation*](http://karpathy.github.io/2015/05/21/rnn-effectiveness/)) ([*code*](https://github.com/karpathy/char-rnn)) ([*reference TensorFlow implementation*](https://github.com/sherjilozair/char-rnn-tensorflow)) in TensorFlow, using methods we’ve learned in class and some additional methods as well (dynamic\_rnn, BasicLSTMCell, and MultiRNNCell). We’ll train it on a corpus of poetry about Providence and other cities, which we are in the process of assembling. Our goal is to assemble at least a couple of megabytes of poetry, which might be challenging, since book-length poetry (especially about Providence) is hard to come by.

**Goals**

*Image Captioning:*

* Assemble data
* Reimplement **im2txt**
* Train

*Language Modeling:*

* Assemble poetry
* Reimplement **char-rnn**
* Train

*Other:*

* Get the Google Maps API pipeline in place
* [Possible] Write a script to run the models and overlay the output poetry on top of the input image.

**Architecture Diagram**

+--------+

| |

+----------> | im2txt | +-------+

| | | |

+-----------+ | +--------+ | +----------+

| | | | | |

| Google | +----+ +----------+ +------> poem +----> | char-rnn |

| Maps | | | | | seed | |

| API | | | Spatial | | +----------+

| | +----------> | Envelope | +-----+

+-----------+ | |

+----------+

**Relevance and Applicability**

This project is not meant to contribute to the field’s state-of-the-art or to apply a pre-existing model to solve an important open problem, but rather to be a (hopefully) enjoyable artistic project that will blend together the feelings that different people have had about Providence and other cities, and echo them in response to real scenes in Providence. We’re hoping that the results will be both sensible and well-formed, and also thought-provoking and maybe even critical. This project will also help us learn more about deep learning, especially encoder-decoder networks and sampling from generative models, in a more relaxed and self-directed setting. We’re looking forward to presenting some results!