**Proposal for CIE6023(MDS6232) Final Project 2018**

**Name:** LIN Jinghong (218012024) SHI Siyuan(218012030) WEI Qing(218012040)

**Title:** Image to Styled Story Telling

**TA-in-charge:**

**Paper ID:**

**Description:**

**Task:**  Generate short stories of particular style basing on images.

**Previous work:**

Karpathy, A., & Li, F. F. (2015). Deep visual-semantic alignments for generating image descriptions. Computer Vision and Pattern Recognition(pp.3128-3137). IEEE.

Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., & Schwenk, H., et al. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. Computer Science.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. (2015, June). Show, attend and tell: Neural image caption generation with visual attention. In International conference on machine learning (pp. 2048-2057).

Chollampatt, S., & Ng, H. T. (2018). A multilayer convolutional encoder-decoder neural network for grammatical Error Correction

**Drawback:**

grammatical error correction;

vapidity of sentence

**Innovation**：

* Generate a paragraph basing on images instead of a single sentence
* Add style translation to the image caption model
* Use Grammatical Error Correction to improve the grammatical performance of the results

**Learning system:**

* Use RNN Encoder–Decoder to train a visual-semantic embedding between COCO images and captions. After training, we can embed new images and retrieve captions.
* Use RNN decoder on Shakespeare's Sonnets. Each passage from the dataset is mapped to a word embedding. The RNN then conditions on the word embedding and aims to generate the passage that it has encoded.
* Bridge the gap between retrieved image captions and passages in novels. Suppose we have 3 vectors: an image caption x, a "caption style" vector c and a "book style" vector b. Then we define F as F(x) = x - c + b, which intuitively means: keep the "thought" of the caption, but replace the image caption style with that of a story. Then, we simply feed F(x) to the decoder.

**Datasets**：

MS COCO, Shakespeare's Sonnets <http://www.gutenberg.org/cache/epub/1041/pg1041.txt>

**To-do lists**:

* Nov 10 - Nov 17: Paper searching, reading, building proper code environment
* Nov 18 - Nov 24: Building up codebase and get down to datasets
* Nov 25 - Nov 30: Implement ideas
* Dec 01 - Dec 07: Fine tune the model
* Dec 08 - Dec 15: paper writing, presentation preparation