

Using Neural Networks to Generate Image Captions

-

— Talk outline —

Tudor Avram

11 Feb 2017

1. Intro to Neural Networks — 15 min

- (a) *Deep Neural Networks* — 7 min
 - i. What is a DNN? — 1 min
 - ii. Single Neuron Model — 2 min
 - Weighted sum
 - output = $P(1|parameters)$
 - iii. Backpropagation — 4 mins
 - 2 phases of the algorithm:
 - A. Propagation
 - B. Weights update
- (b) *Convolutional Neural Networks* — 6 min
 - i. Comparison with DNNs — 2 min
 - ii. What is convolution? — 2 min
 - Parallel to signal processing
 - iii. Convolution layer — 1 min
 - iv. MaxPooling layer — 1 min
- (c) *Recap on NNs* — 2 min
 - i. General Recap — 1 min
 - ii. Application example — 1 min
 - see what-dog.net

2. Generating Captions from images – 13 min

- (a) *Brief overview of the entire system* – 1 min
- (b) *Computer Vision CNN (Generating labels)* – 4 min
 - i. Why CNN? – 1 min
 - Explain why CNN is the best solution for image processing

- ii. CNN model for getting 1 label – 2 min
 - iii. Extending 2.a.ii to multi-labeling – 1 min
 - (c) *Using RNN to generate the captions – 7 min*
 - i. What is an RNN? – 1 min
 - ii. Brief description of the system – 2 min
 - iii. LSTM-based Sentence Generator
 - A. What is LSTM? – 1 min
 - B. LSTM memory cell – 3 min
 - (d) *Overall Loss Function – 1 min*
3. **Demo – 2 min**