Reading your paper:  
first pass

Submitted by Ekin Akyürek

# OVERVIEW & PURPOSE

Reading a technical paper is unlike reading fiction: prepare to spend a lot more time, do multiple passes, take notes, survey literature, figure out the math, understand the algorithms. You are just starting machine learning, so many concepts will be unfamiliar. However you will have a whole semester with your paper and we will follow a process in which your understanding will improve gradually.

# HOW TO READ

To help you with this task please first read a couple of guides on how to read:

* [How to read a paper](http://www.sigcomm.org/sites/default/files/ccr/papers/2007/July/1273445-1273458.pdf) by S. Keshav.
* [How to read a technical paper](https://www.cs.jhu.edu/~jason/advice/how-to-read-a-paper.html) by J. Eisner.

# FIRST PASS

Briefly scan your paper using Keshav’s first pass. Please give the reference to your paper and your one paragraph summary below. The paper reference should have the format “Author, Year, Title, Howpublished” (you can optionally include a link). The summary should be what your current understanding is, not just a paraphrase of the abstract:

**Paper: Donahue, Jeffrey, et al. "Long-term recurrent convolutional networks for visual recognition and description." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.**

**Summary: Convolutional networks is used for object recognition tasks. It can also be used for better feature representation of images. We can embed these features to LSTM networks to get arbitrary length input description of images. Overall model cascades CNN to LSTM decoder and train both network at the same time. Thus, we can get image captions(arbitrary length output) from CNN representation of images(static input). Suggested model uses pre trained CNN networks which previously used for object recognition task and stacks it to 2-layer LSTM network and try to maximize probability of (image,y1,y1,y3..,yt|V,W). X is image, y1,y2… is caption words, V is the convolutional network parameters W is the parameters of LSTM network.**

# UNKNOWN TERMS

Now go through the paper a second time with a highlighter (or the equivalent software tool) and mark all the terms and phrases that you do not fully understand. Make a list of these terms below. I will use this list to make sure we will cover these concepts in class, and assume that you understand everything that is not on this list:

* LSTMs and how we train them -> [9],[10],[13] citations
* Semantic video role tuple predictors -> [11] citation in the article
* Conditional Random Field -> It is used in video-translational part of the article.(not urgent to learn)

# PAST WORK

Go over the paper’s related work section one more time. What are the three most important papers cited by this paper in your opinion? Please give references to these papers and your opinion about why these papers are the most important.

* I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in NIPS, 2014.
  + To understand LSTMs and sequence to sequence learning better
* K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in ICLR, 2015.
  + To understand image.net classification better. To learn image.net datasets and use them. I need to implement pretrained VggNet, CaffeNet and AlexNet convolutional networks.
* T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, “Microsoft COCO: Common objects in ´ context,” arXiv preprint arXiv:1405.0312, Tech. Rep., 2014.
  + One of the new datasets for image labeling is Microsoft’s COCO, I think I should learn how to download and use it.

# RELATED/FUTURE WORK

Use Google Scholar to find out who cited your paper. If you can find your paper in Google Scholar use the “Cited by”, and “Related Articles” links. If you cannot find your paper see if you can find the prior work you cited in the previous section, or if you can find related papers using keywords. Note the citation counts of your results, these give you an indication of popularity if the paper is not very new. What are the three most important follow-up papers in this area in your opinion? Please give references to these papers and your opinion about why these papers are the most important.

* Citation1:Xu, Kelvin, et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention." *ICML*. Vol. 14. 2015.
  + Show Attend and Tell one of the state of art article in the image captioning. It uses similar methods with my article. I can use it to understand cnn-rnn networks better. It is highly cited too.
* Citation2:Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
  + This article is suggests more complicated network, but it is similar to my article in a sense that it combines convolutional networks with rnns. It is highly cited too.
* Citation3:Mao, Junhua, et al. "Deep captioning with multimodal recurrent neural networks (m-rnn)." *arXiv preprint arXiv:1412.6632* (2014).
  + M-RNN is another popular type of network which is used in image captioning. It is somehow more complicated but I think benchmark results are similar. I may use this article to enhance my network in the end.