Emoji-Language Image Captioning

Team: Ian Knight, Rayne Hernandez, Quint Underwood

CS 231N: Convolutional Neural Networks for Visual Recognition

Problem Description

- By what method can we assign emojis that best describe the semantic content of an image?
- By what method can we translate English text captions into emojis?
- How does one embed emojis into the same space as word embeddings?
- What neural architecture best serves the task of emoji-language image captioning?

Dataset

Our dataset consists of the 2017 COCO dataset, which includes images (118K train, 5K test) and their corresponding English-language captions (5 per image) for a total of approximately 715K (image + caption) data points.

Approach

Preprocessing

Part-of-speech filtration

Vectorize captions (word2vec)

Transform to emoji2vec space

Cosine similarity

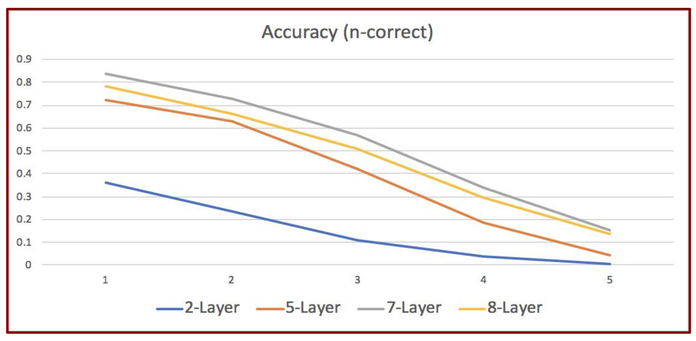
Model Architecture

Resize Image → **CNN (7-conv)** → **3-layer Neural Net** → **Class Softmax**

Emoji Class Selection

- Only use emojis with at least 100 instances in training set captions → 208 emoji classes

Results



Example Results

Conclusion

- We achieved considerable success in generating semantically accurate image captions
- Some overfitting occurs with the 8-conv CNN; the 7-conv CNN provided the best results
- Quantifiable metrics (e.g. recall) alone fail to take account of partially correct predictions
- There is room for improvement in devising an even better method of assigning true emoji labels to images

Citations

[1] Ben Eisner, Tim Rocktaschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. ¨ emoji2vec: Learning emoji representations from their description. CoRR, abs/1609.08359, 2016.
[2]] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4):652–663, April 2017.
[3] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 26, pages 3111–3119. Curran Associates, Inc., 2013.