Team: Two Headed Monster

Facebook Project: No

Project Title: Neural Image Captioning

Project Summary:

One of the fundamental problems in artificial intelligence that connects computer vision and natural language processing is to automatically describe the content of an image. This task involves analyzing an image's visual content and generating a textual description that gives the most salient aspects of an image. Following the work of [1], [2], and [3], neural image captioning—the use of deep neural networks to accomplish this task—has produced the most impressive results, and the use of pretrained image models in an encoder-decoder architecture enables transfer learning for optimal performance. This project aims to explore recent approaches for image captioning and investigate which architectures are best suited to the task and how they can be improved.

Approach:

- Construct a baseline model using a CNN encoder and LSTM decoder based on [2]. Likewise, we will use the Flickr8K, Flickr30K, and COCO datasets for comparison.
- Explore sub-model architectures to improve model performance.
- Explore the use of pre-trained image models with fine tuning to further improve performance.
  - VGG19
  - InceptionV3
  - Resnet-50
  - EfficientNet
- Explore the use of pre-trained language models with fine tuning to further improve performance. (stretch goal)
  - GPT-2
  - Transformer-XL
  - Reformer
  - XLNet

Resources/Related Work:

[1] "Show and Tell: A Neural Image Caption Generator", Vinyals et al.

[2] "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", Xu et al.

[3] "Vivo: Surpassing Human Performance in Novel Object Captioning with Visual Vocabulary Pre-Training", Lin et al.

[4] “Long-term Recurrent Convolutional Networks for Visual Recognition and Description”, Donahue et al.

[5] “What is the Role of Recurrent Neural Networks (RNNs) in an Image Caption Generator?”, Tanti et al.

[6] “Automatic Description Generation from Images: A Survey of Models, Datasets, and Evaluation Measures” , Bernardi et al.

[7] “Where to put the Image in an Image Caption Generator”, Tanti et al.

[8] “Text Augmentation Using BERT for Image Captioning”, Atliha and Šešok

[9] “Vivo: Surpassing Human Performance in Novel Object Captioning with Visual Vocabulary Pre-Training”, Lin et al.

Datasets:

- https://cocodataset.org/#captions-2015
- http://bryanplummer.com/Flickr30kEntities/
- https://www.kaggle.com/adityajn105/flickr8k/activity

Team Members:

Lujia Zhang

Lukas Olson

Xinyi Chen

Jeongho Jo

Looking for more members: No

Pizza Link: https://piazza.com/class/kni1g8lh43l3ie?cid=223_f23