Automatic Image Captioning

Project Proposal

The objective of the project is to create an automatic Image Captioning model that takes in an image as input and outputs a sequence of text that describes the image. In order to achieve this we would be making use of CNN(Convolutional Neural Network) and RNN(Recurrent Neural Network). The applications of automatic image captioning are as listed below:

* Useful for blind people who rely on sound or text to understand a scene.
* Making the web content accessible.
* Describing a video in real time.

We would be using the COCO(Common Objects and Contexts) dataset to train and test the neural network. A CNN preserves spatial information and images while an RNN works well with sequential data, such as generating word sequences. The images are fed into the CNN. Furthermore, we’ll be using a pre-trained network like VGG16 or Resnet.