

PROJECT

Submitted by : Athirakrishna P. R

Topic: Automatic Image Captioning Using Deep Learning

Guide: Ms. Remya P.S., MTech, MISTE , Asst Professor, CSE Dept

About project:

Interpreting and abridging the bits of knowledge picked up from pictures is a tedious undertaking. As a result, there is an impressive requirement for programmed strategies that can rough the mapping from visual data to literary portrayals. To accomplish the objective of image captioning, semantic information of pictures should be caught and described in natural language. Image captioning is a very difficult undertaking, because it requires interfacing both research areas of NLP (natural language processing) and computer vision. Here using an end-to-end neural network system that can automatically view an image and generate a reasonable description in plain English. It is based on a convolution neural network that encodes an image into a compact representation, followed by a recurrent neural network that generates a corresponding sentence.

In this project I use a neural and probabilistic framework to generate descriptions from images. These models make use of a recurrent neural network which encodes the variable length input into a fixed dimensional vector, and uses this representation to “decode” it to the desired output sentence. Thus, it is natural to use the same approach where, given an image (instead of an input sentence in the source language), one applies the same principle of “translating” it into its description. Here replacing the encoder RNN by a deep convolution neural network (CNN). Over the last few years it has been convincingly shown that CNNs can produce a rich representation of the input image by embedding it to a fixed-length vector, such that this representation can be used for a variety of vision tasks.

The task of image captioning can be divided into two modules logically- one is an image based model – which extracts the features out of image, and the other is a language based model – which translates the features and objects given by image based to a natural sentence. For image based model (encoder) this system rely on convolutional neural network model. And for language model (decoder) I use a long short term memory network.

PUBLICATIONS

- Paper - Study on pedestrian detection
Conference - ICICT 2018
Journal - IE E E
Status – Presented
- Paper - Performance study of object detection models
Conference - ICE CCT 2019
Journal - IE E E
Status - Accepted