Image Captioning

Sharhad Bashar | 260519664
Professor: Dr. Jeremy Cooperstock | Graduate Student: Roger Girgis
McGill University | Faculty of Electrical and Computer Engineering | ECSE 499 Honors Thesis

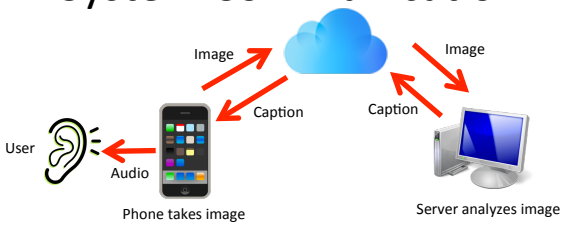
Autour

Overview

This project is focused on developing an image captioning system for Autour. Autour is an eyes-free mobile system developed to aid the visually impaired get a better understanding of their surroundings. This thesis discusses the implementation of Deep Neural Networks that analyze an image and generates a caption describing the image as well as any text that may be present.

User with Autour

System Communication



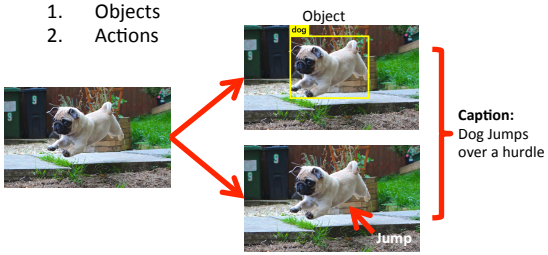
Motivation

PASCAL Visual Classes and Imagenet ILSVCR:

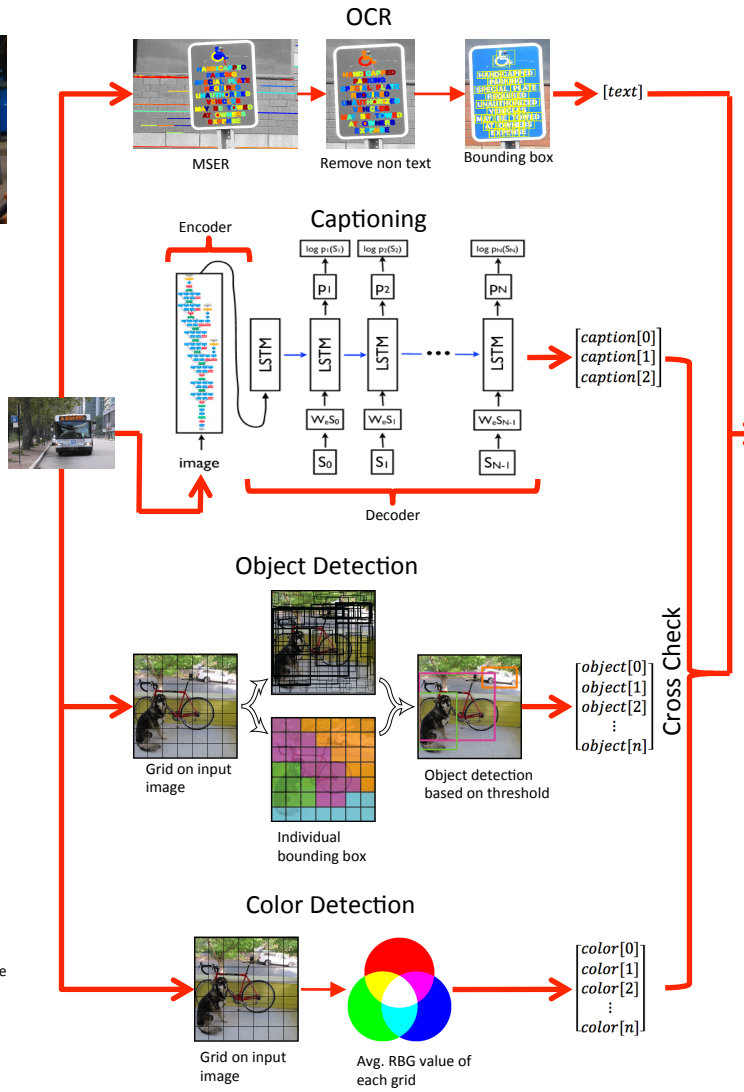
1. Create dataset of images and descriptions
2. Contest to test new algorithms and models

Ingredients of an image:

1. Objects
2. Actions



Model Implementation



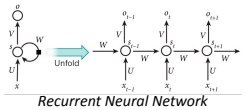
Final Caption:
A white bus driving down a street next to tall buildings
Text reads: MERCY BERGAN

Model Components

OCR (Optical Character Recognition)

1. Maximally Stable External Regions (MSER)
   - Uniform Intensity
   - Surrounded by contrasting backgrounds
2. Remove non text region
3. Create bounding boxes
4. Combine bounding boxes
5. Apply OCR to get the text

Captioning



Recurrent Neural Network

Input Image I
Maximize P(S|I)

1. Encoder:
   - 48 Layer CNN
   - Input: Image
   - Output: Vector rep. of image
2. Decoder:
   - RNN with LSTM blocks
   - Input: Vector rep. of image
   - Output: 3 captions with confidence

Object Detection

1. 32 Layer CNN
2. Break image into small squares
3. Generate Bounding boxes based on confidence
4. Combine bounding Boxes
5. Apply threshold

Color Detection

1. Break image into small squares
2. Avg. RGB value of each square
3. Use it to get a range of 11 most common colors:

Cross Check

1. Generate a list of objects and their synonyms
2. Generate a list of colors
3. Count the number of objects and colors in caption
4. Caption with highest count and confidence score chosen