

## **CS6301 MACHINE LEARNING – MINI PROJECT PROPOSAL**

SRIHARI. S

T.K.S. ARUNACHALAM

2018103601

2018103616

### **IMAGE INSCRIPTION AND INTONATOR - A NEURAL NETWORK APPROACH**

---

Technology has become an integrated part of our daily lives over the past decades. The efficient processing and association of different multimodal information is a very important research field with a great variety of applications, such as human computer interaction, knowledge discovery, document understanding, etc. Computerized elucidation of an image has been one of the primary goals of computer vision. Not only must description generator models be powerful enough to solve the computer vision challenges of determining which objects are in an image, but they must also be capable of capturing and expressing their relationships in a natural language. It is a very important challenge for machine learning algorithms, as it amounts to mimicking the remarkable human ability to compress huge amounts of salient visual information into descriptive language.

Hence, to tackle this conundrum we present the development of a novel methodology to extract meaningful information from images, in the form of short descriptions. The results can be further run through a lexical articulator engine to offer full sustainability. This way, a full independent experience could be delivered to visually impaired people.

For an accurate algorithm, it must be fed with good amount of data. The dataset which we are planning to use for training and testing the model is MS-COCO (Microsoft-Common Objects in Context) which contains more than 80000 labelled images.

For converting still images into natural language text sentences we must first start from the understanding of the context of an image and secondly how this context is expressed into natural language. Thus, for understanding the image a feature extracting convolutional neural network can be used. With the aid of recurrent neural networks, the extracted features can be transformed into a suitable textual description. These results are finally fed to a lexical articulator module which produces the output of the system in the form of speech.

#### **References:**

- A. Puscasiu, A. Fanca, D. Gota and H. Valean, "Automated image captioning," 2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), Cluj-Napoca, Romania, 2020, pp. 1-6, doi: 10.1109/AQTR49680.2020.9129930.
- P. Jayawardhana, A. Aponso, N. Krishnarajah and A. Rathnayake, "An Intelligent Approach of Text-To-Speech Synthesizers for English and Sinhala Languages," 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 2019, pp. 229-234, doi: 10.1109/INFOCT.2019.8711051.