# MINI **PROJECT** PROPOSAL

|  |  |
| --- | --- |
| NAME | MRIDUL GUPTA PARTH GUPTA |
| ROLL NO | 2K17/SE/63 2K17/SE/77 |
| DEPARTMENT | SOFTWARE SOFTWARE  ENGINEERING ENGINEERING |
| PROJECT TITLE | IMAGE CAPTIONING |
| FACULTY ADVISOR | Dr. RUCHIKA MALHOTRA |
| AIM OF THE PROJECT | **This project aims at generating captions for images. The project has its base on Neural Networks, which uses a variant of Recurrent neural Network (RNN) coupled with a Convolutional Neural Networks (CNN). We intend to use phrases as elementary units instead of words, which would lead to better semantic and syntactical captions. The project further can be extended to generate audio for the text generated.** |

# ABSTRACT AND SPECIFIC AIMS

Image captioning refers to the problem of constructing natural language description of an image. This is an important problem with practical significance that involves two major artificial intelligence domains- Computer vision and natural language processing. The architecture combines image feature information from a Convolutional Neural Network (CNN) with a Recurrent Neural Network (RNN) language model, in order to produce sentence-length descriptions of images. We evaluate the proposed model on publicly available benchmark Flicker8k dataset.

Auto captioning could, for example, be used to provide descriptions of website content, or to generate frame-by-frame descriptions of video for vision impaired.

Platforms like facebook can infer directly from the image where you are (Beach, Café etc), what you wear (color) and more importantly what you are doing (in a way).

## INTRODUCTION

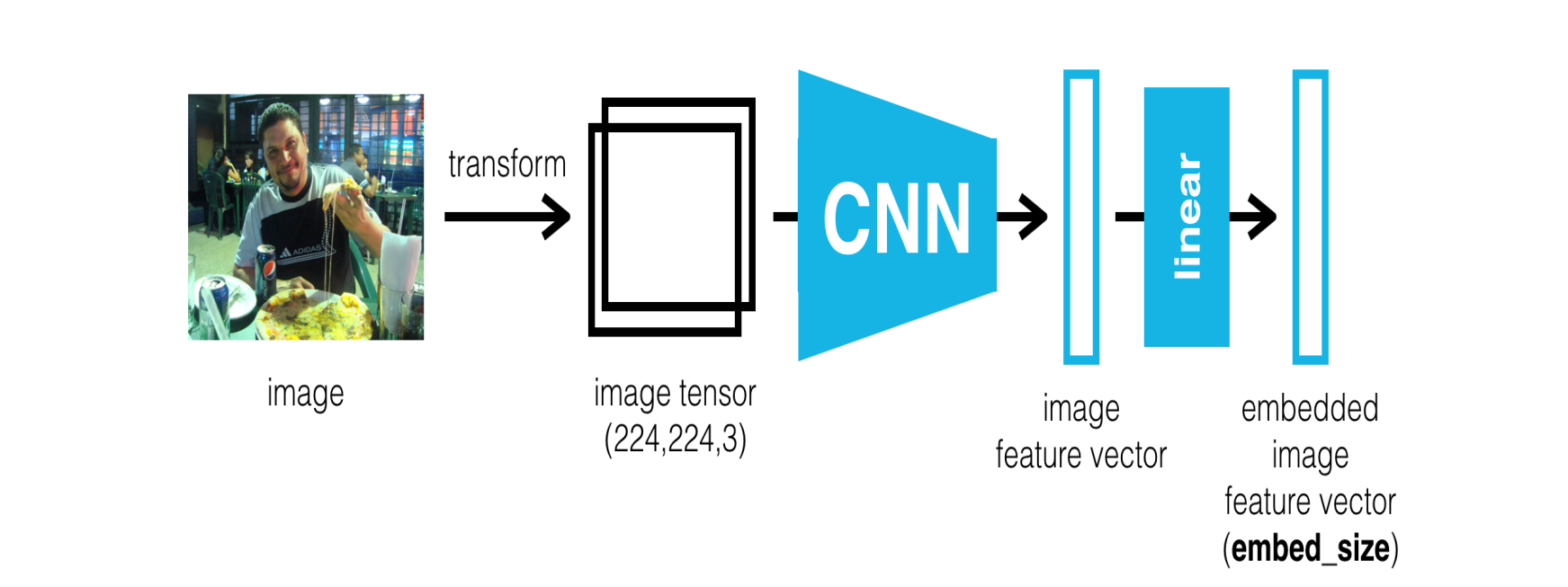
Automatic Image captioning refers to the problem of constructing a natural language description of an image. This task is challenging than the image classification and object recognition task, because it not only requires detection of objects within the image, but also requires detection of their relationship, expression and activity presented in the image. Furthermore, the perceived information must be translated to some human understandable natural language (like English). An important application of image captioning system is in aiding visually impaired persons by providing then information about the content of the image in natural language. Another application is in search engines where images can be searched by sentence fragments.

Key steps of image captioning task include extracting salient high level features from an image, detecting objects from those features, detecting salient casual information (relationship interaction, expression, activity) involving those objects, and finally generating a natural language description as a sequence of words to express the content off an image.

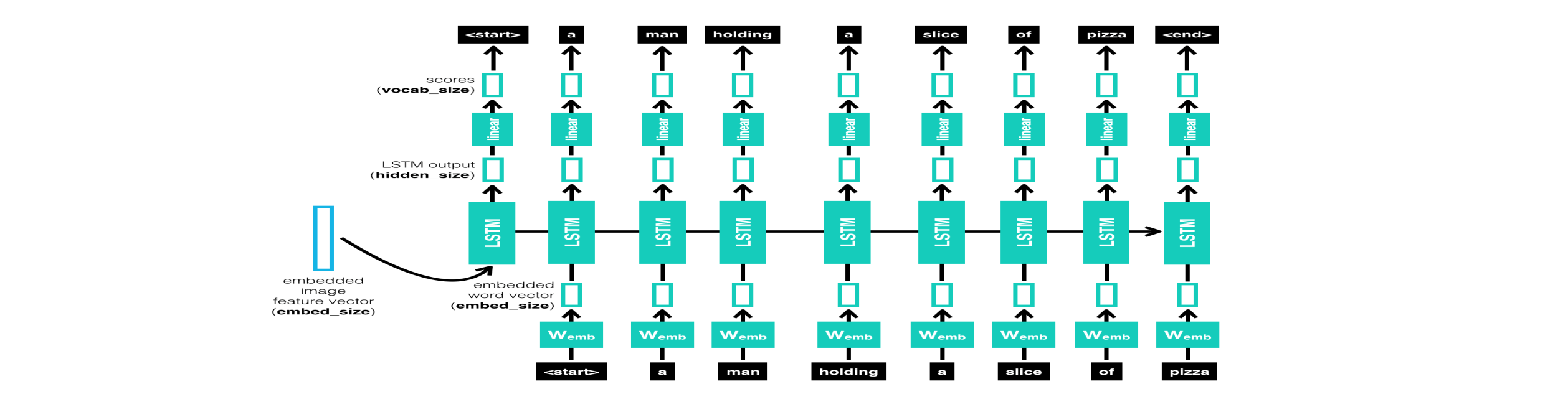
# PROJECT DESIGN

## Network Topology

**Encoder:**   
We use Convolutional Neural Network (CNN) as our encoder. The image is given to CNN to extract the relevant features. The last hidden state in CNN is connected to Decoder The encoder that we provide to you uses the pre-trained VGG architecture (with the final fully-connected layer removed) to extract features from a batch of pre-processed images. The output is then flattened to a vector, before being passed through a Linear layer to transform the feature vector to have the same size as the word embedding.

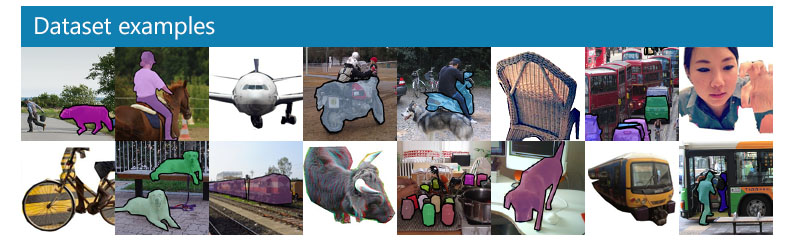
[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/encoder.png)

**Decoder:**   
We use Recurrent Neural Network (RNN) as our encoder which it takes the features from encoder and produces a sentence for it.

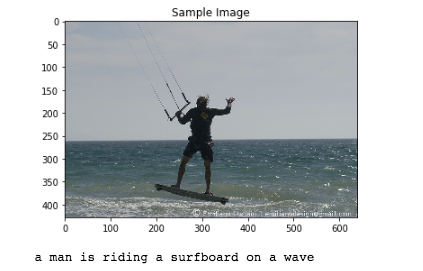
[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/decoder.png)

## DATASET

Flickr8k dataset is a large-scale dataset for scene understanding. The dataset is commonly used to train and benchmark object detection, segmentation, and captioning algorithms.

[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/coco-examples.jpg)

Expected Results:

[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/sample_3.png)[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/sample_1.png)