



# **LOW LEVEL DOCUMENT**

**SUBMITTED BY**

**PRATEEK MAGDUM**

(IMAGE CAPTIONING PROJECT)

---

**INEURON INTERNSHIP**

---

Domain : Security and Safety, Surveillance

**IMAGE - CAPTIONING**

# Contents

## **1 Introduction**

1.1 Why this Low-Level design document?

1.2 Scope

1.3 Constraints

## **2 Technical Specifications**

2.1 Introduction

2.2 Motivation & Challenges

2.3 Procedure

## **3 Results**

3.1 Required Results

3.2 Outcomes

# 1 Introduction

## 1.1 Why this Low-Level design document?

The purpose of this LLD or a Low-Level Design (LLD) document is to give the internal logical design of the actual program code for Image captioning project. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document. This document is intended for both the stakeholders and the developers of this project and will be proposed to the higher management for its approval.

The main objective of the project is to analyse the various aspects with different use cases which covers many aspects of Image captioning . It helps in not only understanding the meaningful relationships between attributes but it also allows us to do our own research and come-up with our findings.

## 1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

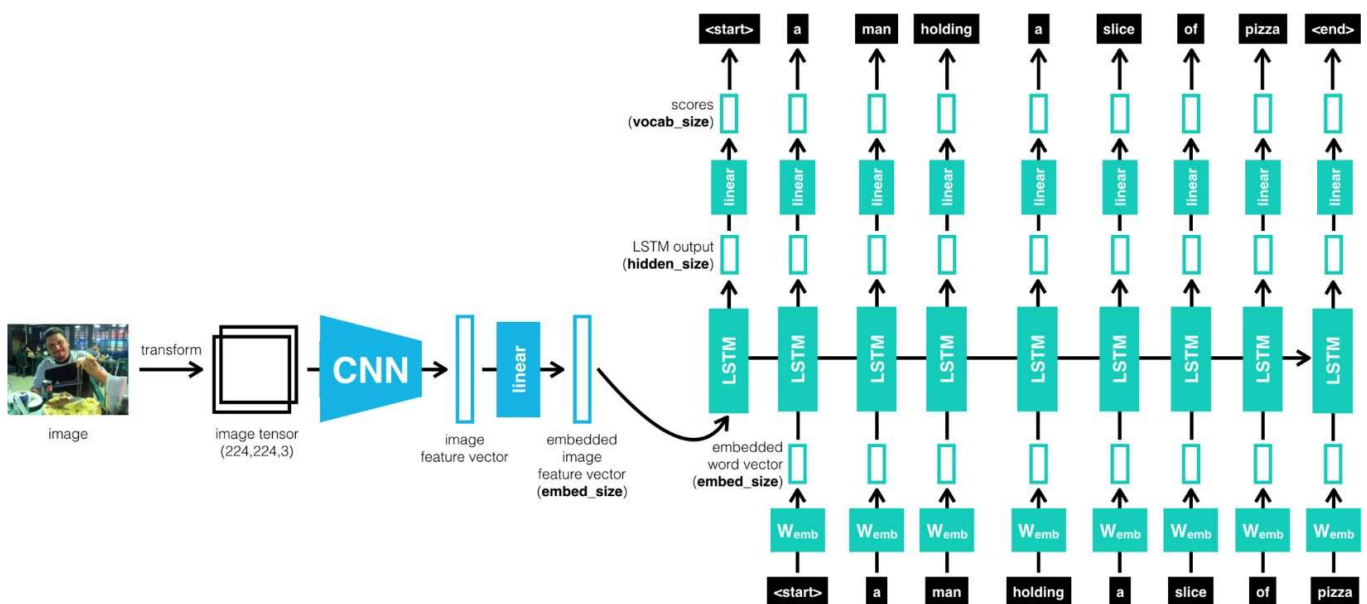
This study demonstrates the how different analysis help out to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

## 1.3 Constraints

The analysis must be user friendly, code must be neat & clean, EDA must be automated as much as possible because it will save huge amount of time. Moreover, users should not be required to have any of the coding knowledge as the insights they are looking for are mentioned in-detail with respective visuals.

## 2.1 INTRODUCTION

• **Image-Captioning** We, as a race, are in an era where social media has completely enveloped our day-to-day lives. Major players in this sector deal with digital data in the size of billions. To connect with the world and tell their stories, people upload pictures with captions. They come with catchy phrases to draw people to their profiles. But, what if we could automate this captioning process; what if we could take an image, study it using an algorithm and generate relevant headings and captions? This would not only enhance the way we use social media, but it would transform the way we look at images in general. Our project in deep learning aims at helping with this and serving as a template for various sectors of society to benefit from its applications.



## 2.2 MOTIVATION & CHALLENGES:

**Motivation/Purpose:** Motivation: Image captioning can not only enhance the way social media is used, it can also help in improving lives of communities, especially the visually impaired. Developing applications for them would help them listen to images. In addition to that, image captioning can also play a vital role in the healthcare industry where it can support medical image reading and finding anomalies in them to help physicians better. All-in-all, image captioning could revolutionize the way we look at images.

**Challenges:** One of the challenges is image semantics richness where the images used for testing must be semantically related to those used for training the model. Inconsistent objects during training and testing is another hurdle in which the existing training process relies heavily on the selection of data sets. Cross-language text description of images is also an issue and how to implement cross language text description of images is a key problem and a research difficulty in image captioning.

## 2.3 PROCEDURE

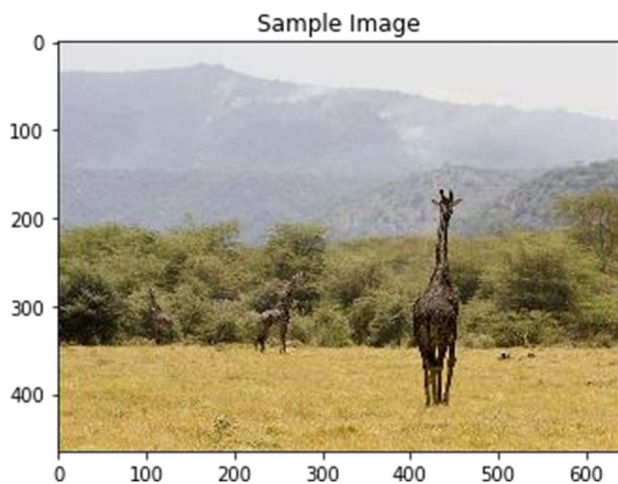
**Techniques used:** We are planning on using the Flickr 8K dataset for our project. Considering that we have two diverse sets of features, images and captions (text), we would be using a combination of architectures that can encompass the features present in the data. The techniques used are: 1. Convolutional Neural Network (CNN): Using pre trained-ImageNet CNN model to obtain image feature vectors. 2. Word Vectors: Build index dictionary to encode word to integers using GloVe word2vec model. 3. LSTM: Since we are approaching it as a supervised learning problem, we would be using partials captions with image vectors to train the LSTM model to get target captions.

## 3. RESULTS

### 3.1 REQUIRED RESULTS

#### - Some Visualization

Input Image



**Caption:**

A jiraffe Standing in a field

Input Image



**Caption:**

A man takes a slice of pizza from its plate.

Input Image



**Caption:**

A surfer is riding a wave in the ocean

## 3.2 EXPECTED OUTCOMES

Expected Outcomes: The basic objective of image captioning is to describe an image in text format and hence that would be the major outcome of the project. The resultant captions would cover major image properties like objects, activities, counts and colours. Theoretical and practical implications: Image Captioning has been widely used in different forms ranging from image-indexing for visually-impaired individuals to group images on platforms like Google and Facebook. The idea can be extended to videocaptioning, audio-video annotation, and the popular Visual-Question Answering (VQA) challenge which becomes more user-interactive as we ask questions regarding the input image and receive answers in various format