**PROGRESS REPORT**

**PROJECT: IMAGE CAPTIONING**

|  |  |
| --- | --- |
| 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

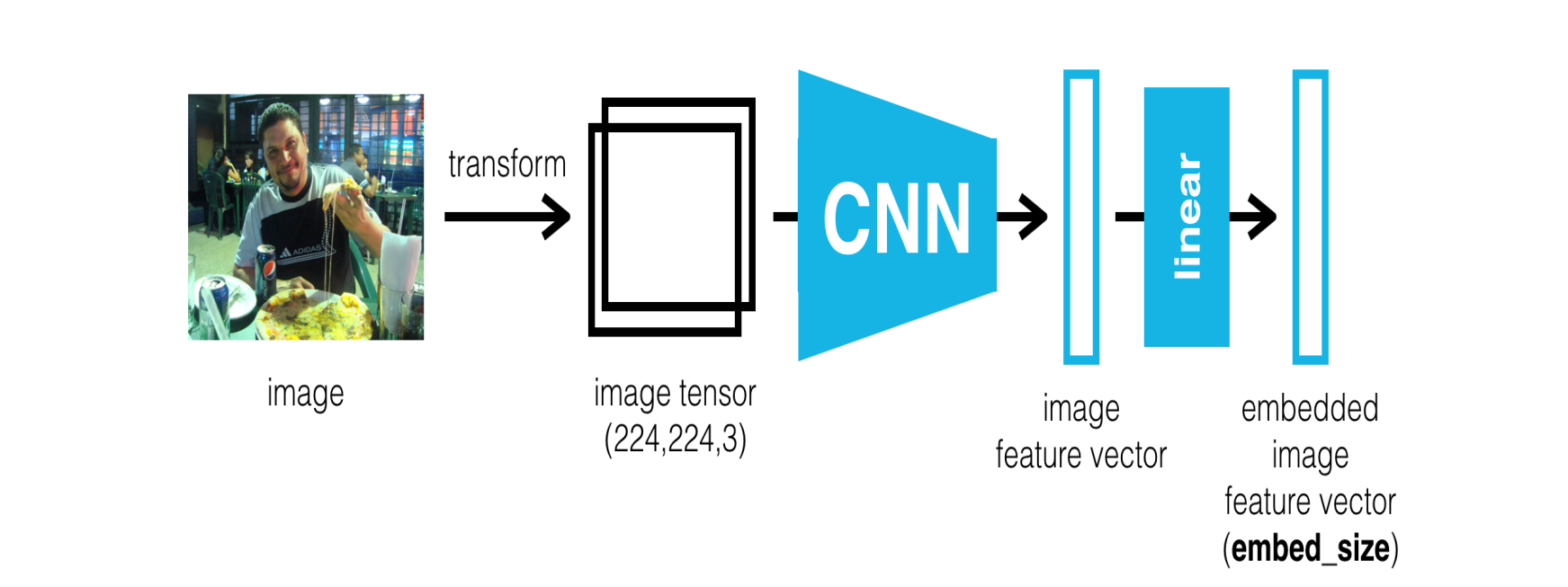
Automatic Image captioning refers to the problem of constructing a natural language description of an image. This task is challenging than the image classification and object recognition task, because it not only requires detection of objects within the image, but also requires detection of their relationship, 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.

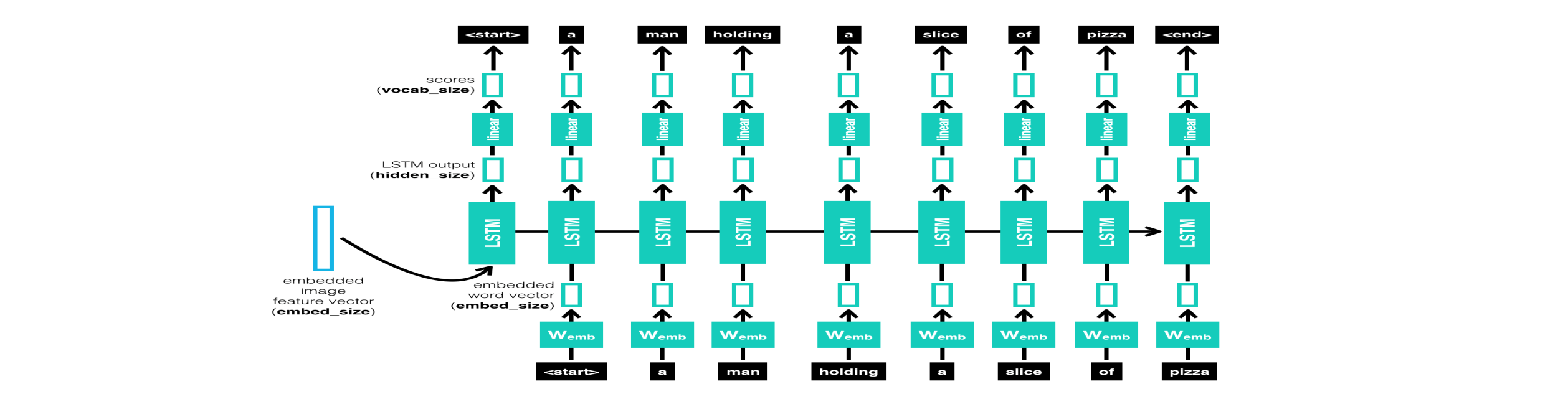
# PROJECT DESIGN

## 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.

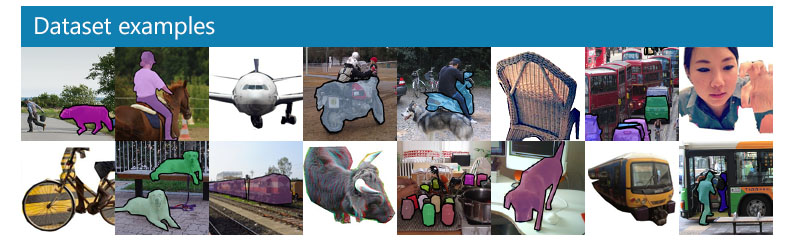
[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/encoder.png)

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

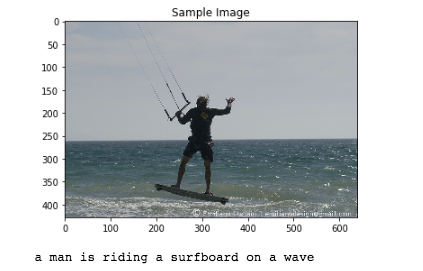
[](https://github.com/soheillll/Automatic-Image-Captioning/blob/master/Images/decoder.png)

## 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)

SOFTWARE REQUIREMENT SPECIFICATION DOCUMENT

FOR

IMAGE CAPTIONING

|  |
| --- |
| 1. Introduction    1. Purpose    2. Scope    3. Definitions, Acronyms, and Abbreviations    4. References    5. Overview |
| 1. Overall Description    1. Product Perspective       1. System Interfaces       2. User Interfaces       3. Hardware Interfaces       4. Software Interfaces       5. Communication Interfaces       6. Memory Constraints       7. Operations       8. Site Adaptation Requirements    2. Product Functions    3. User Characteristics    4. Constraints    5. Assumptions and Dependencies    6. Apportioning of Requirements |
| 1. Specific Requirements    1. External Interface    2. Functions    3. Performance Requirements    4. Logical Database Requirements    5. Design Constraints       1. Standards Compliance    6. Software System Attributes       1. Reliability 3.6.2Availability 3.6.2Security       2. Maintainability       3. Portability    7. Organizing the specific requirements       1. System Mode       2. User Class 3.7.3Objects 3.7.4Feature 3.7.5Stimulus 3.7.6Response   3.7.7 Functional Hierarchy  3.8 Additional Components |
| 4 Supporting Information |

**INTRODUCTION**

This document aims at defining the overall software requirements for “Image Captioning System”. Efforts have been made to define the requirements exhaustively and accurately. The final product will be having only features/functionalities mentioned in this document and assumptions for any additional functionality/feature should not be made but any of the parties involved in developing/testing/implementing/using this product. In case it is required to have some additional features, a formal change request will need to be raised and subsequently a new release of this document and /or product will be produced.

* 1. Purpose

This specification document describes the capabilities that will be provided be the software application. It also sates the various required constraints by which the system will abide. The intended audience for this document is development team, testing team and end uses of the product.

* 1. Scope

The name of the application is Image Captioning System wna dwill be referred to as ICS for the rest of the SRS. The proposed ICS will be able to perform:

DO’s:

1. Issue of login id and password to system operators.
2. Maintain Photo and their captions within the system.
3. Allow the user to rate and review their experience.

Don’ts:

1. The system won’t allow the user to share the generated captions via other applications.
2. Multiple Photographs can’t be uploaded at a time.

Benefits:

1. The system involves minimal manual intervention in processing of data and therefore be free of errors.
2. It would allow efficient generation of description of Images provided.
   1. Definitions, Acronyms and abbreviation
3. SRS- Software Requirement Specification
4. DBA- Database Administrator
5. ICS- Image captioning system
6. Image ID: It is a unique sequence number allocated to each image being processed.
7. System Administrator: User having all the privileges to each photo.
   1. References

1. IEEE Recommended Practice for Software Requirements Specification- IEEE Stud 830-1998

2. Object oriented software Engineering – Yogesh Singh and Ruchika Malhotra.

1.5 Overview

The reset of this SRS document describes the various system requirements, interfaces, features and functionalities in detail.

1. Overall Description

Image captioning refers to constructing a natural language description of an image. The perceived information must be translated to some human understandable natural language (like English).

This software includes tasks like 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.

The admin/Manager will have to maintain the following information:

1. Image database
2. User Database

The admin will have the following functions:

1. Maintain application
2. Provide regular reports

The user will have the following functions:

1. Login
2. Upload image
3. Delete Image
4. Review System

Product functions

The application will provide access to users who have a profile, after that users will be able to perform various functions.

The major functions of the application are –

1. A login facility for enabling only authorized access to the system.
2. The admin will be able access the image database and user database.
3. The user will be able to upload image, delete image and review the application.

User Characteristics

• Qualification: At least senior secondary; should be comfortable with English language

• Experience: Should be well versed about the basic outline of the application.

• Technical Expertise: Should be comfortable using general purpose applications on a computer.

Constraints

1. There will be only one administrator.

2. The user will not be allowed to update the primary key.

3. Due to limited features of the DBMS database auditing will not be provided

4. Administrator must implement a security policy to safeguard the information of hotel guests and rooms from being manipulated by unauthorized users.

Assumptions and dependencies

1. The details related to the user profile will be provided by user.
2. Administration is created in system already.
3. The details of images will be available to admin.

Functional Requirements

Login

|  |
| --- |
| Introduction - This use case documents the steps that must be followed for a user to log in to the application. |
| Actors – Admin, User |
| Pre-Conditions  All users must have a User Account (i.e., User ID, Password) created for them by administrator prior to executing the use case. |
| Post Conditions  If the use case is successful, the Administrator/Hotel Guest/Manager is logged into the system. If not, the system state is unchanged. |
| Event flow  Basic Flow   1. The system requests that the actor enter his login id, password. The role of the user is determined based on the login id. 2. The system validates the entered credentials and logs the actor into the system |
| Alternate Flow  Alternate Flow 1: Invalid Login Details  If in the basic flow, the actor enters an invalid login id, password the system displays an error message. The user can choose to either return to the beginning of the basic flow or cancel the login at which point the Use Case ends.  Alternate Flow 2: User Exits  This allows the user to exit at any time during the use case. The use case ends |
| Special Requirements : None |

Validity Checks

1. Every user will have a unique user id
2. User ID can’t be blank
3. User ID can have only 4-15 characters
4. User ID will not accept blank spaces and special characters
5. Password can’t be blank
6. Length of password can only be from 4-15 characters
7. Alphanumeric characters, hyphen and underscore are allowed in the password field
8. Password will not accept blank spaces

Upload Image

|  |
| --- |
| Introduction: This use case documents the steps that must be followed to upload an image in the system |
| Actors: Admin, User |
| Pre-Conditions The User must be logged into the system before this use case begins. |
| Post Conditions If the use case is successful, the image is uploaded in the system. If not, the system state is unchanged. |
| Event Flow  Basic Flow:  This use case starts when the User wishes to upload an image in the system.  1. The user selects the image to be uploaded.  2. The system validates and image is uploaded. |
| Alternate Flow 1: Image is already uploaded  If the image is already uploaded, the system displays an error message, at which point the use case ends.  Alternate Flow 2: Image deleted  If at any point of time, the User decides to cancel the upload, the use case ends. |
| Special Requirements  None |
| Associated Use Case  Login |

Validity Checks

1. Only the Admin has the permission to access the image database
2. Every uploaded image will have a unique ID
3. there should be an uploaded image.

Error Handling/Response to Abnormal situation

If any of the validation/sequencing flows does not hold true, appropriate error message will be prompted to the administrator to do the needful

Delete image

|  |
| --- |
| Introduction: This use case documents the steps that must be followed to delete an image in the System |
| Actors: Admin, User |
| Pre-Conditions  The user must be logged into the system before this use case begins |
| Post Conditions  If the use case is successful, the image is deleted. If not, the system state is unchanged. |
| Event Flow  Basic Flow:  This use case starts when the user wishes to delete an image from image  1. The system validates the request and image is deleted. |
| Alternate Flow 1:  User Exits  If at any point of time, user decides to cancel the delete request, the use case ends. |
| Special Requirements  None |
| Associated Use Case  Login, upload image |

Error Handing/Response to Abnormal Situations

If any of the validation/sequencing flows does not hold true, appropriate error message will be prompted to do the needful

Caption rating

|  |
| --- |
| Introduction: This use case documents the steps that must be followed for a user to rate the generated caption |
| Actors: user |
| Pre-condition: The user must be logged into the system. |
| Post-condition: Rating is saved in the database. |
| Basic flow:  1. The user successfully logs into the system.  2. The user gives his/her ratings in the column specified for the same purpose.  3. Submit button is clicked and the post is saved in database. |
| Alternate flow 1: User Exits  1. The user decides to cancel the post and exists from the system. The use case ends |
| Special Requirements: None |
| Associated use cases: Login, upload image |

Validity Checks

* The guest must give a rating between 1-5 stars

Error Handling/Response to abnormal situations

If any of the validation/sequencing flows does not hold true, appropriate error message will be prompted to do the needful

REQUIREMENTS:

* THE MODEL SHOULD BE ABLE TO GENERATE SUITABLE CAPTION DESCRIBING THE IMAGE.
* **WE INTEND TO USE PHRASES AS ELEMENTARY UNITS INSTEAD OF WORDS, WHICH WOULD LEAD TO BETTER SEMANTIC AND SYNTACTICAL CAPTIONS.**
* **AN INTERFACE (WEB OR ANDROID) SHALL BE MADE FOR USER INTERACTION.**

**SUMMARY**

* **SO FAR, THE MODEL IS ABLE TO GENERATE A CAPTION FOR THE IMAGE.**
* **WE HAVE BEEN ABLE TO ACHIEVE TARGET OF COMPLETING THE PROJECT UPTO 30%.**
* **DEVELOPMENT OF SRS DOCUMENT HAS BEEN STARTED.**

**TO BE DONE IN FUTURE:**

* **COMPLETION OF SRS DOCUMENT.**
* **IMPROVEMENT IN ACCURACY OF THE MODEL IN GENERATING THE CAPTIONS.**
* **INTERFACE (WEB OR ANDROID) IS TO BE BUILT.**