

Image Captioning

Presented by *Rishabh Chaurasia, Pranav Goyal, Utkarsh Kuchhal*

Under the guidance and support of our mentor: **Vasu Gupta**

Content

Overview

Project objective

Preprocessing Applied

Libraries & Platforms Used

Model Architecture

Model Summary

Predictions

Dataset - Flickr8k

Project timeline- 3 weeks

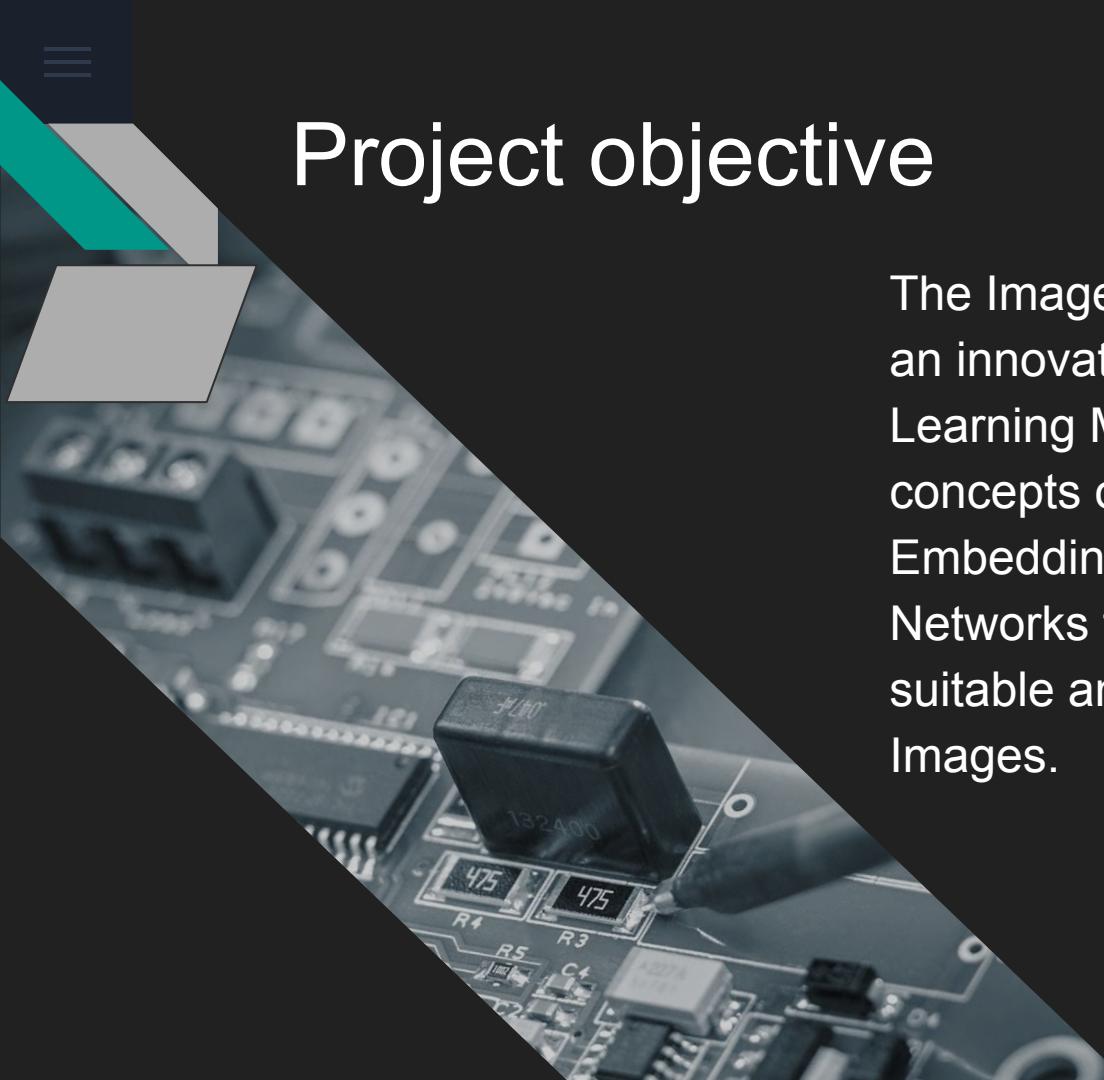
Overview

Automatic Image Captioning

Automatic image annotation is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database.

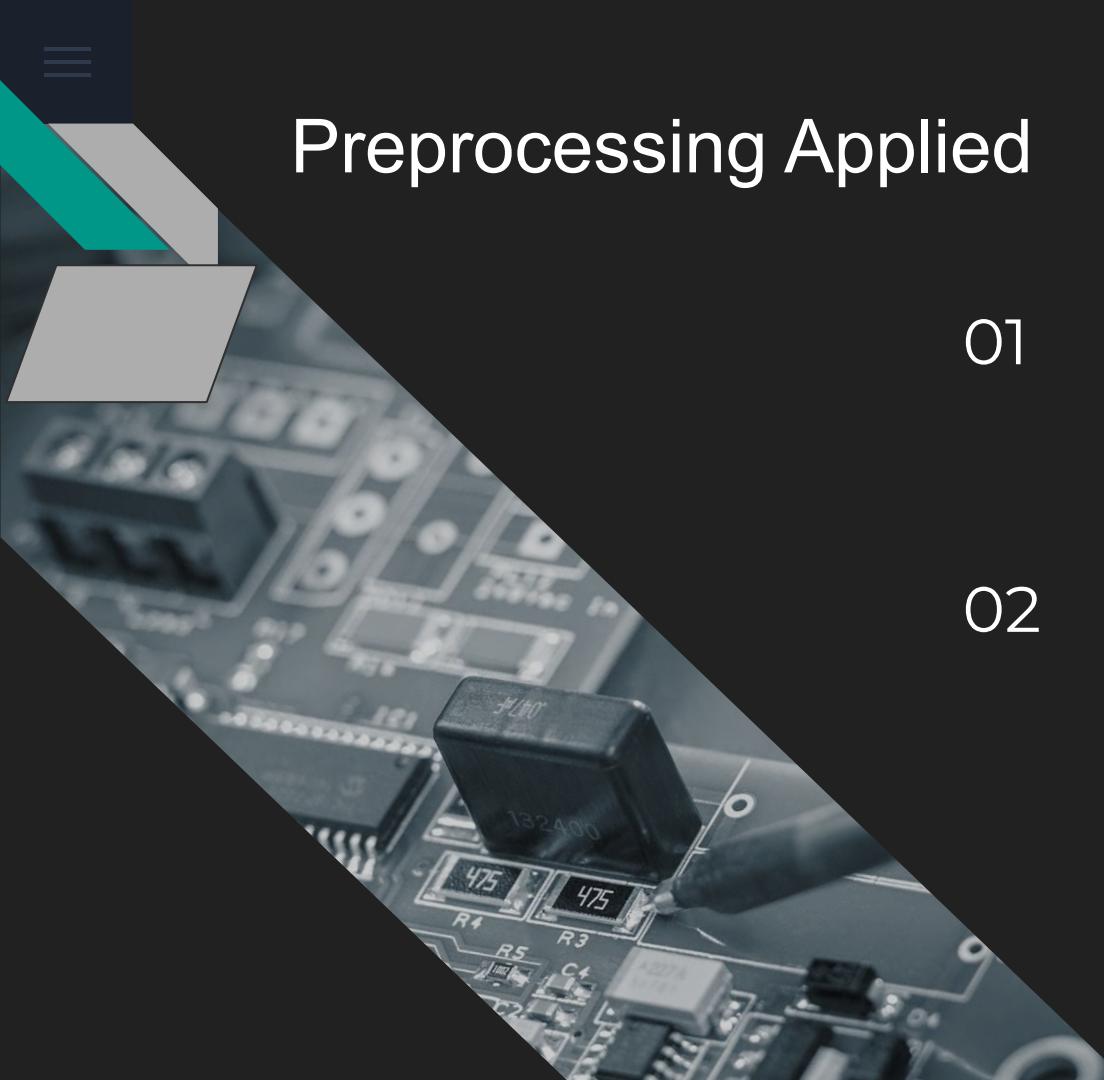
Photo Captions

Photo captions, also known as cutlines, are a few lines of text used to explain and elaborate on published photographs.



Project objective

The Image Captioning project deals is an innovative and enhanced Deep Learning Model that utilises the concepts of Computer Vision, Word Embedding and Convolutional Neural Networks to automatically generate suitable and apt captions for input Images.



Preprocessing Applied

01

Image Preprocessing

02

Text Preprocessing

Libraries and Platforms Used

The entire model was trained and coded in Python3 on Jupyter and Google Collab Notebooks for the efficient compilation of executable codes on various code cells in a simple and user-friendly manner.



Libraries & Platforms Used

The following are the list of libraries used to facilitate the functioning of our code:

1. NumPy - Python library used for working with arrays, linear algebra, fourier transform, and matrices.
2. Matplotlib - Plotting library which provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits.
3. KERAS - Open Source Neural Network library written in Python that runs on top of Theano or Tensorflow, designed to be modular, fast and easy to use
4. Time module - provides many ways of representing time in code, such as objects, numbers, and strings
5. Python pickle module - used for serializing and de-serializing a Python object structure.
6. Pandas - a high-level data manipulation tool, built on the Numpy package and its key data structure is called the DataFrame.
7. Natural Language Processing module(NLTK) - guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more.
8. Collections module- was introduced to improve the functionalities of the built-in collection containers.Collections in Python are containers that are used to store collections of data.

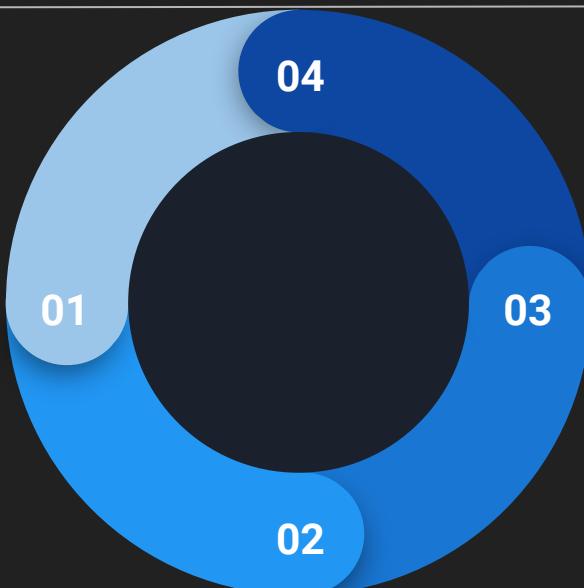


Model Architecture

Image
Encoding with
ResNet

With Input resized to
224X224X3

Word
Embedding of
entire
vocabulary
Using Glove Embedding



Predictions

Outputs the final caption and the
input image for a better
understanding of the user.

Generator
Function

Predicts and Generates word for
the caption one by one based on
the image encoding passed to
the trained model.

Model Summary

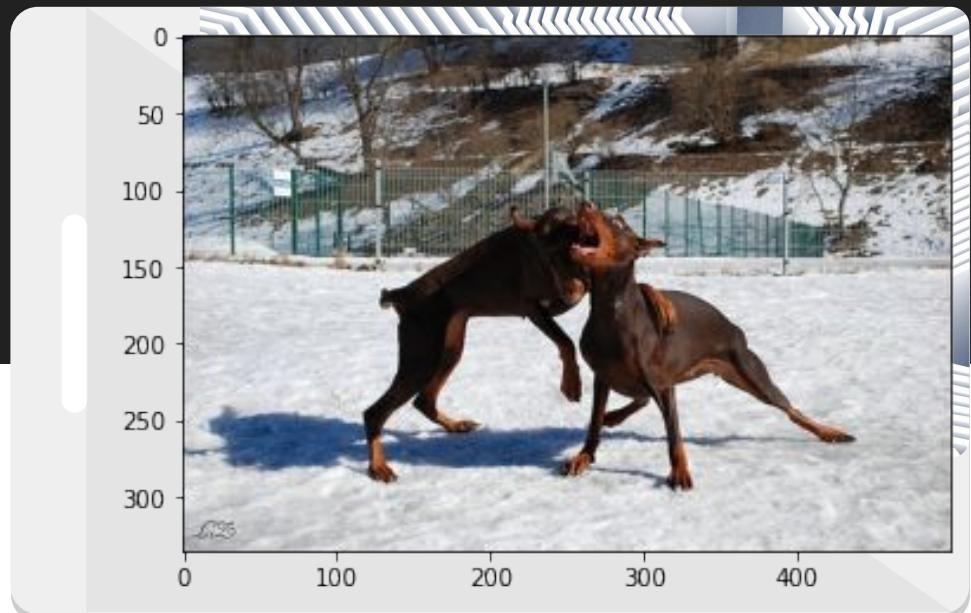
```
model.summary()
```

```
Model: "functional_3"
```

Layer (type)	Output Shape	Param #	Connected to
input_4 (InputLayer)	[None, 35]	0	
input_3 (InputLayer)	[None, 2048]	0	
embedding (Embedding)	(None, 35, 50)	92400	input_4[0][0]
dropout_1 (Dropout)	(None, 2048)	0	input_3[0][0]
dropout_2 (Dropout)	(None, 35, 50)	0	embedding[0][0]
dense_1 (Dense)	(None, 256)	524544	dropout_1[0][0]
lstm (LSTM)	(None, 256)	314368	dropout_2[0][0]
add (Add)	(None, 256)	0	dense_1[0][0] lstm[0][0]
dense_2 (Dense)	(None, 256)	65792	add[0][0]
dense_3 (Dense)	(None, 1848)	474936	dense_2[0][0]

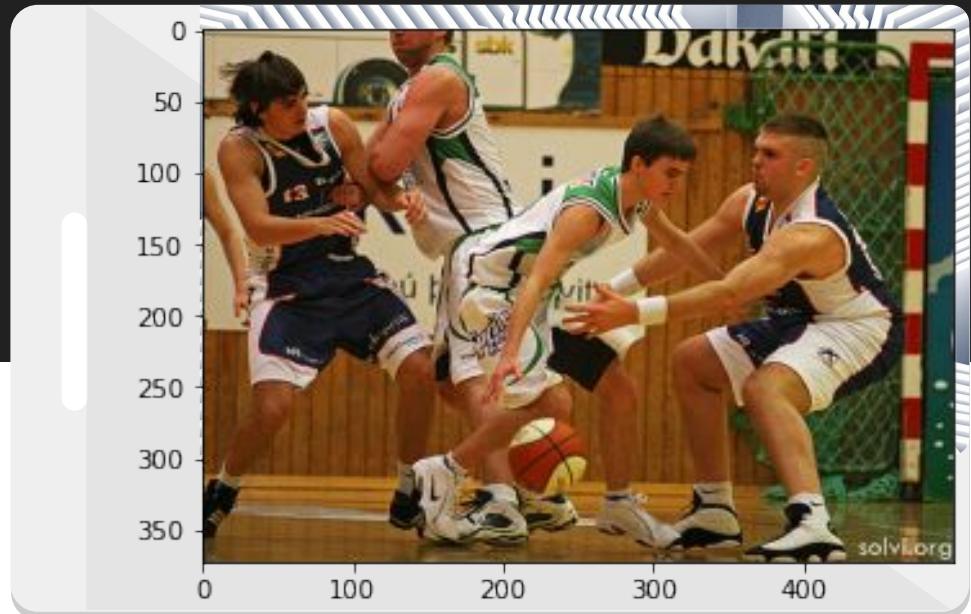
Predictions #1

two dogs are playing in the snow



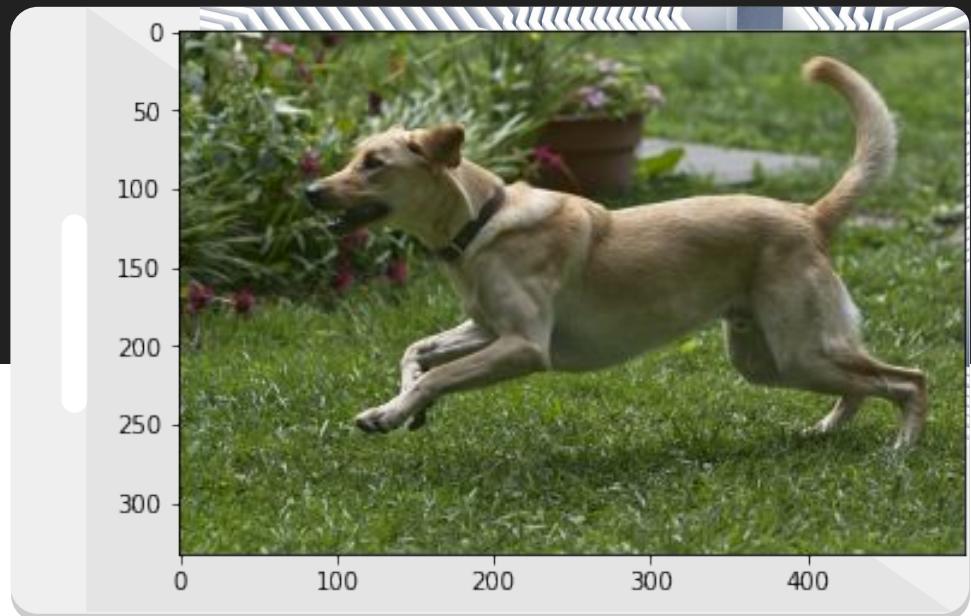
Predictions #2

basketball player in white uniform is trying to make basket but of the player in white tries to stop him



Predictions #3

dog is running through backyard



Thank you!

