Title: Image Captioning using deep learning

Team members
1. 1602-20-737-121
2. 1602-20-737-314

Abstract

This project improves image captioning by combining object detection with traditional methods. It uses pre-trained models for image features and attention mechanisms to focus on important parts, resulting in more accurate and descriptive captions.

Guide _Signature:

VASAVI COLLEGE OF ENGINEERING

PROJECT WORK ALLOTMENT FOR B.E. VIII-SEMESTER STUDENTS

S No.	Roll No.	Student Name	Project Title	Name of the Guide	Place of working the project (Dept. or outside)*
1.	737-121	N. Abhiram	Image Captiaouing Using Deep Learning	Kezia Ravi	
2.	737-314	S. Sai Venkatesh			

If outside, the name of the organization _____

Project Coordinator

HOD

VASAVI COLLEGE OF ENGINEERING

PROJECT WORK ATTENDANCE REPORT

Name of the Department: **Information Technology (IT)**

Name of the Guide: Internal **Kezia Rani**

External: _____

(if project work carried- out outside)

Project work title: **Image captioning using Deep learning**

Place of working: Department/Name of the Industry/Institution (if outside) _____

S No	Name of the student	Reg. No.	No. of classes attended for the periods							No. of classes attended for all periods (cumulative)
			08/1/24 To 20/1/24	21/1/24 To 03/2/24	4/2/24 To 17/2/24	18/2/24 To 02/3/24	03/3/24 To 16/3/24	17/3/24 To 30/3/24	31/3/24 To Last date of Instruction	
1	Abhiram	602-20 757-12||							
2	Venkatesh	314								
3										
4										
5										

SIGNATURE OF THE GUIDE

SIGNATURE OF PROJECT COORDINATOR

SIGNATURE OF THE HOD

Progress Report during the period 04/02/24 To 17/02/24 (2 weeks)

Project work title: _____ Image Captioning Using Deep learning

S No.	Progress Report	Signature of the student
1.	using the collected dataset we started the implementation of the project	

Team Members : __N. Abhiram, S. Saivenkatesh__

Roll Numbers : __1602-20-737-121, 1602-20-737-314__

Remarks of the Guide

Coding is in process. Suggested them to complete coding & write documentation also parallelly.

Signature of the Internal Guide

Signature of the Coordinator

Signature of the HOD

Image Captioning Using Deep Learning

C.S.Kanimozhiselvi
Professor,
Department of Information Technology,
Kongu Engineering College,
Erode, Tamilnadu, India.
kanimozhiselvi.cse@kongu.edu

Karthika V
UG Scholar,
Department of Computer Science and Engineering,
Kongu Engineering College, India
karthihkavenkatachlalam2000@gmail.com

Kalaivani S P
UG Scholar,
Department of Computer Science and Engineering,
Kongu Engineering College, India.
kalaivanisp2000@gmail.com

Krithika S
UG Scholar,
Department of Computer Science and Engineering,
Kongu Engineering College, India.
keerthisomu8@gmail.com

Abstract

The process of generating a textual description for images is known as image captioning. Now a days it is one of the recent and growing research problem. Day by day various solutions are being introduced for solving the problem. Even though, many solutions are already available, a lot of attention is still required for getting better and precise results. So, we came up with the idea of developing a image captioning model using different combinations of Convolutional Neural Network architecture along with Long Short Term Memory in order to get better results. We have used three combination of CNN and LSTM for developing the model.

The proposed model is trained with three Convolutional Neural Network architecture such as Inception-v3, Xception, ResNet50 for feature extraction from the image and Long ShortTerm Memory for generating the relevant captions. Among the three combinations of CNN and LSTM, the best combination is selected based on the accuracy of the model. The model is trained using the Flicker8k dataset.

Keywords— Convolutional Neural Network, Xception, Inception v3, ResNet 50, Long Short Term Memory.

I. INTRODUCTION

The process of creating a textual explanation for a set of images is known as image captioning. The captions are generated using both Computer vision and NLP. Because it converts images from a sequence of pixels to a series of words, image captioning can be viewed of as an end-to-end Sequence to Sequence challenge.Both the language or remarks as well as the images must be processed for this purpose. Long Short Term Memory is employed for the Language component and Convolutional Neural Networks is used for the Image part.

Extraction and Language Model are two features. A phase in the dimensionality reduction process is feature extraction, which divides and reduces a big set of raw data into smaller groupings. As a result, processing will be more simple. As a result, by selecting and merging variables into features, feature extraction aids in the extraction of the best feature from big data sets, effectively reducing the amount of data. A language model is an important part of current Natural Language Processing (NLP). It is a statistical method for predicting words from human language patterns.For captioning image, we are building an LSTM-based model that uses feature vectors from the training network to predict caption sequences.