

Certifiate

This is to certify that this is the bonafide record of the application development entitled, ”<AD Title Name>” submitted by

Achanta Vineesh Chowdary : 2211CS020004

Anguru Karthik : 2211CS020014

B. Keshavardhan Reddy : 2211CS020027

Bollam Srujith : 2211CS020075

B. Bhargava Sai Abhinay : 2211CS020076

of B.Tech III year Ist semester, Department of CSE (AI&ML) during the year 2024- 25.The results embodied in there port have not been submitted to any other university or institute for the awardof any degree or diploma

|  |  |
| --- | --- |
| **INTERNAL GUIDE** | **Head of the Department**  **Dr. A.Sivaranjani** |

|  |  |
| --- | --- |
| **External Examiner** | **DEAN-AIML**  **Dr. Thayyaba Khatoon** |

1. INTRODUCTION
   1. Abstract
   2. Limitations of project
2. LITERATURE REVIEW
3. PROBLEM STATEMENT
4. DATA COLLECTION & EXPLORATION
5. MODEL SELECTION & ARCHITECTURE

INTRODUCTION

* 1. Abstract
  2. Limitations of project

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Paper Title  (2020) | Methodology | Advantage | Disadvantage |
| 01 | Chetan Amritkar, et.  al., 2018 [4] | This model incorporates  both recurrent neural  network (RNN) and  Convolutional Neural  Network (CNN) are used  to extract feature from  image. The model is  trained so that when an  input image is provided as  input it produce the  captions which clearly  describes the image. | The model you describe, which incorporates both a Recurrent Neural Network (RNN) and a Convolutional Neural Network (CNN), leverages the strengths of both architectures to generate image captions. **Feature Extraction (CNN)**: CNNs are excellent at extracting spatial features from images. **Sequential Processing (RNN)**: RNNs, especially in the form of Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU), are powerful for handling sequential data | The model’s descriptions  or captions are divided  into three categories:  description error free,  description containing few  small mistake, image and  description are slightly  linked but, not at all. The  categories in the results  are caused by the  proximity of some specific  words; for example, when  a ‘vehicle’ is nearby, term  like “vehicle”,” Car”,”  Van “etc. are also  formatted, which may be  inaccurate it is evident  from a vast number of  studies that using larger  dataset improves the  model’s performance. |
| 02 | Jyun-You Lin, et.  al., 2020 [5] | Smart glasses system which is Based on deep learning  system for visually  impaired people. By  capturing photo from the  camera function of the  smart glasses, the system  can upload captured  images to our object  detection system which  function at backend and  provide voice speech of  caption helpful to  understand visually  disabled individuals  regarding the object  Infront of them. | A smart glasses system based on deep learning offers transformative benefits for visually impaired individuals by providing real-time object recognition and voice feedback. The system captures images through a built-in camera and uses deep learning models, such as Convolutional Neural Networks (CNNs), to detect objects and describe them through speech. This gives users instant, accurate information about their surroundings, allowing them to navigate independently. By offering continuous feedback, the glasses enable visually impaired individuals to identify objects and understand their spatial arrangement, promoting autonomy in various environments. | smart  glasses system which is  Based on deep learning  system for visually  impaired people. By  capturing photo from the  camera function of the  smart glasses, the system  can upload captured  images to our object  detection system which  function at backend and  provide voice speech of  caption helpful to  understand visually  disabled individuals  regarding the object  Infront of them. |
| 03 | **Raimonda Stani ¯ ut˙e and Dmitrij Šešok** | The SLR has become a great help in the dynamic, data driven world of today, with massive data  volume growth. It is sometimes very di\_cult to consume all currently existing information before  starting to delve into a specific field. In this case, when we talk about image captioning and, as already  said, having so much meaning in this task, it was found that there is much literature, which is hard to  summarize and thus stay up to date with the newest achievements | Sometimes it is defined as a structure of CNN + RNN. Usually a convolutional neural network  (CNN) represents the encoder, and a recurrent neural network (RNN) the decoder. The encoder is the  one which “reads” an image—given an input image, it extracts a high-level feature representation.  The decoder is the one which generates words—given the image representation from the encoder  (encoded image), it generates words to represent the image with a full grammatically and stylistically  correct sentence. | A network is proposed that  varying the guide map  reflecting the style  associated to region of  image, creates multiple  captions. The CAM vector  creates latent space which  we first receive at VAE  encoder the Vector  representing the image  attention region style is  then taken out of the laten  space. This vector serves  as condition to construct  the caption at the VAE  decoder where it is used at |
| 04 | Genc Hoxha, et. al.,  2019 [7] | The RS image retrieval  system have three primary  processes. The first one  involves generating the  Textual description with  Convolution neural  network (CNN) and  produces written  explanations of the  images' contents. In order  to extract features from  images, a convolutional  neural network (CNN) and  a recurrent neural network  (RNN) are combined in  the second step. The third  step involves producing  the descriptions and  details of the content,  respectively. | The RS (Remote Sensing) image retrieval system, which utilizes a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), offers several key advantages through its three-step process. Firstly, by employing CNNs to generate textual descriptions, the system efficiently produces detailed and accurate explanations of image content. This helps in understanding complex remote sensing images, which may include landscapes, urban areas, or other large-scale environments. Secondly, the integration of CNNs and RNNs allows for a comprehensive feature extraction process. | With the goal of  examining the high-level  sematic material buried in  the created descriptions,  we have presented in this  study a sematic picture  retrieval technique based  on generated textual  description. we find that  there is an average  difference of 0.3 in mean  BLEU score when  comparing the produced  description and the  genuine description for RS  pictural retrieval. Our goal  is future work is to  enhance the caption  generating block in order  to close this gap and boost  retrieval capabilities |
|  |  |  |  |  |
|  |  |  |  |  |

LITERATURE REVIEW

PROBLEM STATEMENT

<<--Should articulate the issue that needs addressing

--Frame the problem statement, kind of solutions ,

--Output of the Models ( Ex : Prediction, Detection……)>>>>

DATA COLLECTION & EXPLORATION

--Data Preprocessing techniques

--Gather and curate the necessary data for your project

--Conduct exploratory data analysis to understand the dataset/image set…

--Clean and preprocess the data, handling missing values and outliers

--Perform data transformation and feature engineering as needed

MODEL SELECTION & ARCHITECTURE

(Choose the appropriate deep learning model(s) Design the architecture of your model and its layers )