INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

IMAGE BASED QUESTION AND ANSWERING SYSTEM

[1]Mr B NARSINGAM, [2]B.SUMANTH, [3]D.GANESH CHANDU, [4]K.SANDEEP

[1]Assistant Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

tkrcsebnarsingam@gmail.com

[2,3,4]BTech Student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad

sumanthbelladhi@gmail.com, ganeshchandu29@gmail.com, sandeepkotakonda4648@gmail.com

ABSTRACT: *Image based question answering is a useful way of finding information about physical objects. Current question answering (QA) systems are text-based and can be difficult to use when a question involves an object with distinct visual features. An image QA system allows direct use of a image to refer to the object. We develop a three-layer system architecture for image QA that brings together recent technical achievements in question answering and im- age matching.*

Keywords*: Image based question answering system,*

I. INTRODUCTION

vision-to-Language problems present a particular challenge in Computer Vision because they require translation between two different forms of information. In this sense the problem is similar to that of machine translation between languages. In machine language translation there have been a series of results showing that good performance can be achieved without developing a higher-level model of the state of the world. In [1], for instance, a source sentence is transformed into a fixed-length vector representation by an 'encoder' RNN, which in turn is used as the initial hidden state of a 'decoder' RNN that generates the target sentence. Despite the supposed equivalence between an image and a thousand words, the manner in which information is represented in each data form could hardly be more different. Human language is designed specifically so as to communicate information between humans, whereas even the most carefully composed image is the culmination of a complex set of physical processes over which humans have

ISSN: 2057-5688

little control. Given the differences between these two forms of information, it seems surprising that methods inspired by machine language translation have been so successful. These RNN-based methods which translate directly from image features to text, without developing a high-level model of the state of the world, represent the current state of the art for key Vision-to-Language (V2L) problems, such as image captioning and visual question answering

II. LITERATURE SURVEY

A deep convolutional neural network architecture by the name of Inception is proposed. It establishes a new standard for detection and classification (ILSVRC14). The most basic characteristic of this particular architecture is the increased usage of resources of computing inside the network. Even inside the convolutions, the essential approach to do this is to move from fully connected to sparsely connected structures. State-of-the-art sparse arithmetic operations result from grouping sparse matrices into comparatively dense submatrices. Increasing the scale of deep neural networks is the most simple method for improving their performance. This includes increasing the depth and the breadth of the network. Another advantage of this method is that it is based on the idea of processing the visual input to numerous scales before being aggregated, allowing the following step to abstract data from multiple scales at once.

III. PROPOSED SYSTEM

To bypass the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method where we use selective search to extract just 2000 regions from the image and he called them region proposals. Therefore, now, instead of trying to classify a huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated using the selective search algorithm.

SYSTEM ARCHITECTURE

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

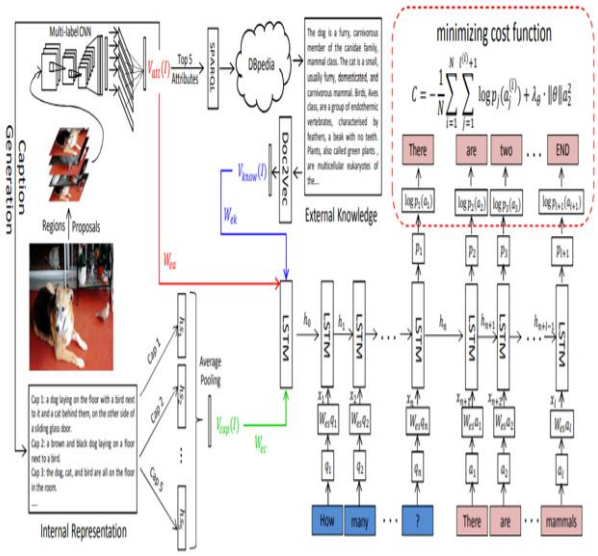


Fig.1 System architecture

Machine Learning :

Before we take a look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models tunable parameters that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain.Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

Categories Of Machine Leaning :-

At the most fundamental level, machine learning can be categorized into two main types: supervised learning and unsupervised learning.

ISSN: 2057-5688

Supervised learning involves somehow modeling the relationship between measured features of data and some label associated with the data; once this model is determined, it can be used to apply labels to new, unknown data. This is further subdivided into classification tasks and regression tasks: in classification, the labels are discrete categories, while in regression, the labels are continuous quantities. We will see examples of both types of supervised learning in the following section.

Unsupervised learning involves modeling the features of a dataset without reference to any label, and is often described as "letting the dataset speak for itself." These models include tasks such as clustering and dimensionality reduction. Clustering algorithms identify distinct groups of data, while dimensionality reduction algorithms search for more succinct representations of the data. We will see examples of both types of unsupervised learning in the following section.

Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven't surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, "to make decisions, based on data, with efficiency and scale".

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programing logic, in the problems that cannot be programmed inherently. The fact is that we can't do without human intelligence, but other aspect is that we

ISSN: 2057-5688

all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises

IV. RESULTS

Image Question Answer

The dataset which you gave us can answer limited questions such as 'does sphere and cube colour is same' or 'sphere is present in image' then it will answer 'YES' or 'No' or number of objects in image but our project can describe anything in the image by using user question and to answer this we are using RCNN to identify objects in images and LSTM to build vocabulary sentences in meaningful form

To run project double click on 'run.bat' file to get below screen

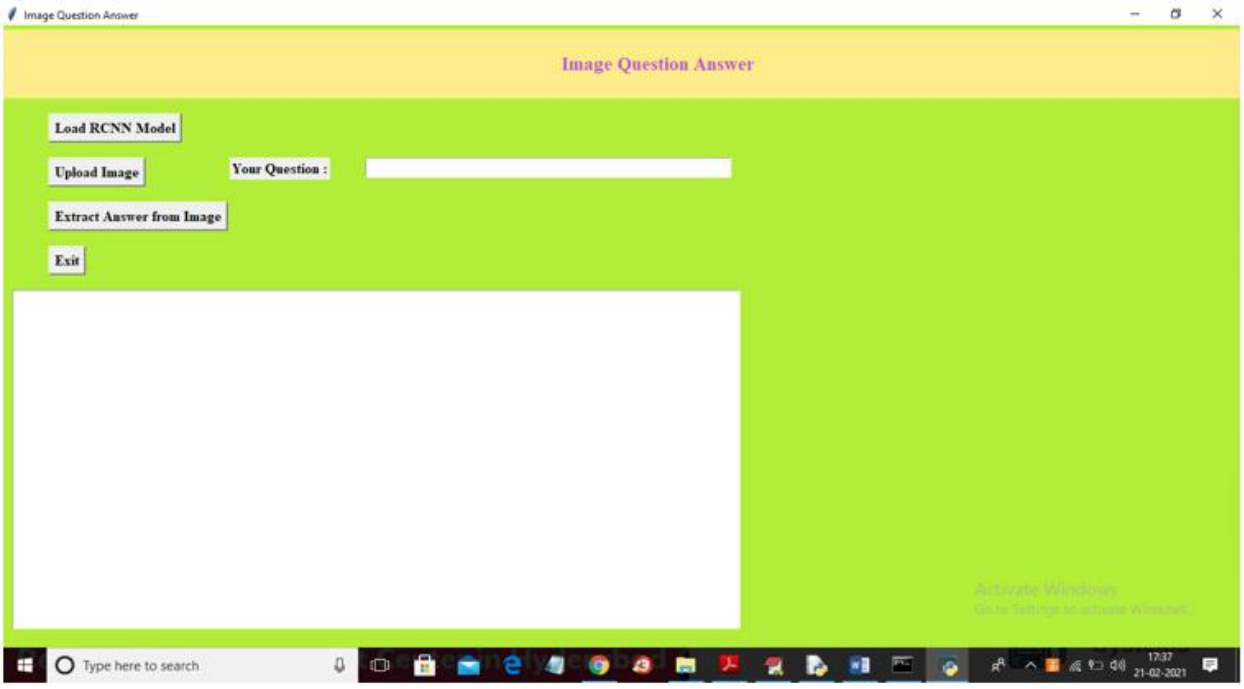


Fig.2 In above screen click on 'Load RCNN Model' button to load CNN model and to get below screen

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

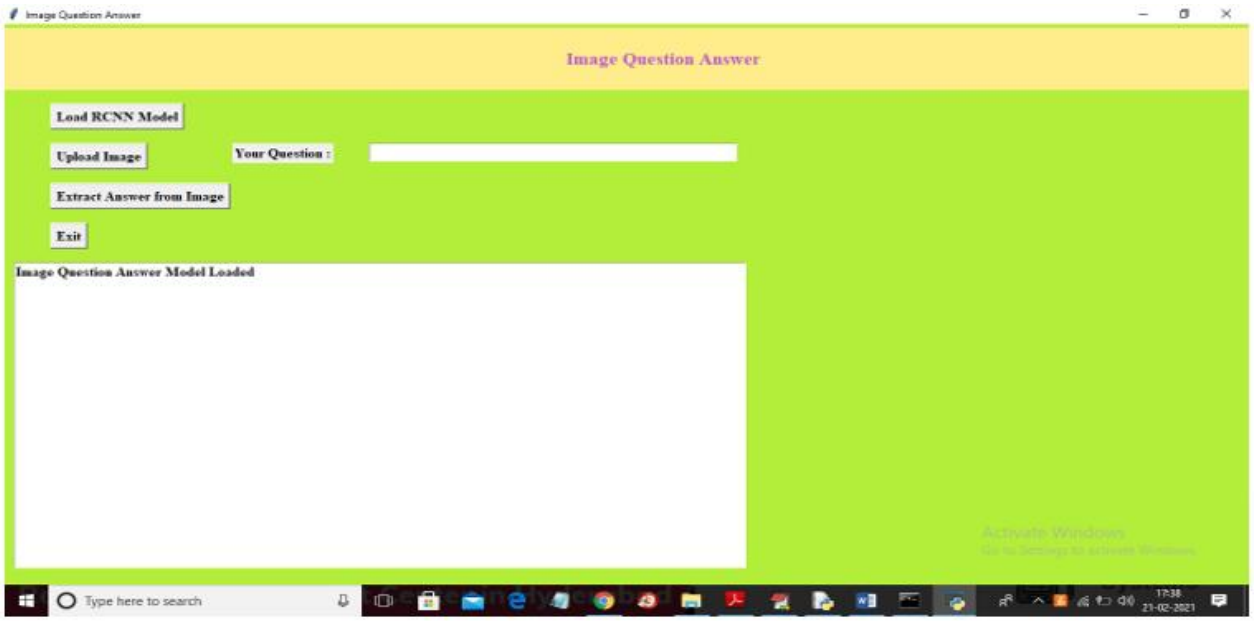
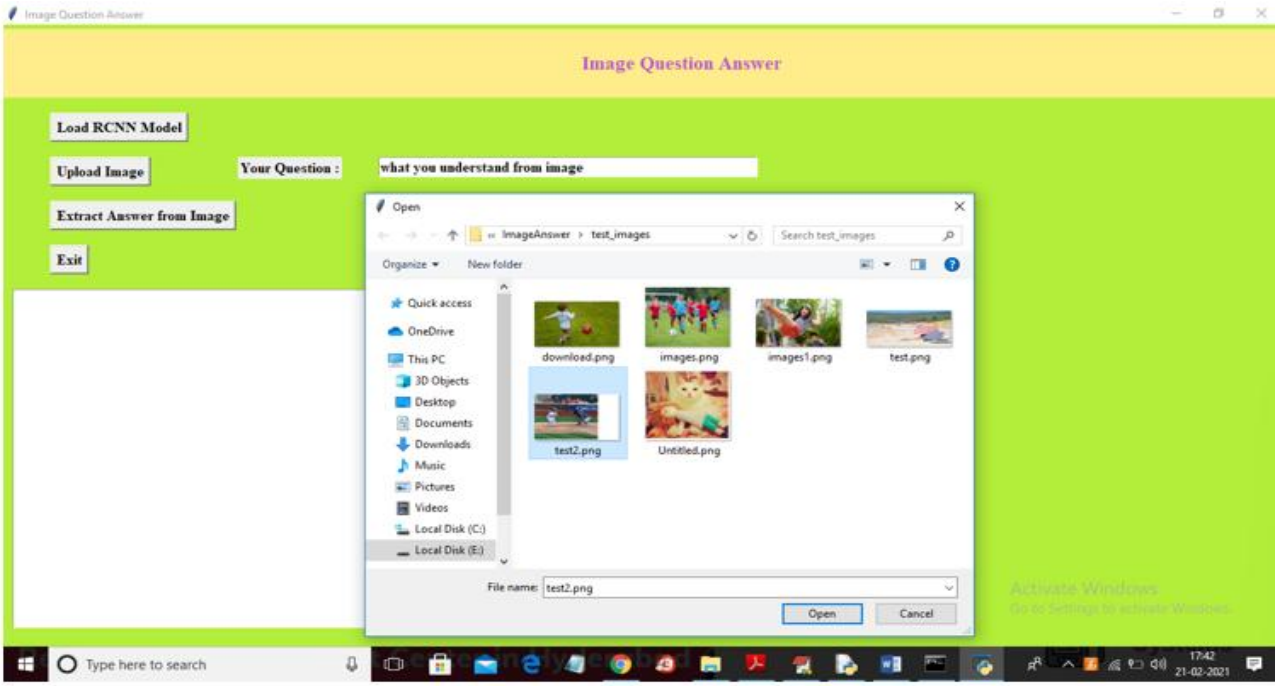


Fig.3 In above screen in text area we can see model loaded and now click on 'Upload Image' button to upload



ISSN: 2057-5688

Fig.4 In above screen selecting and uploading ‘test2.png’ file and then click on ‘Open’ button to get below screen

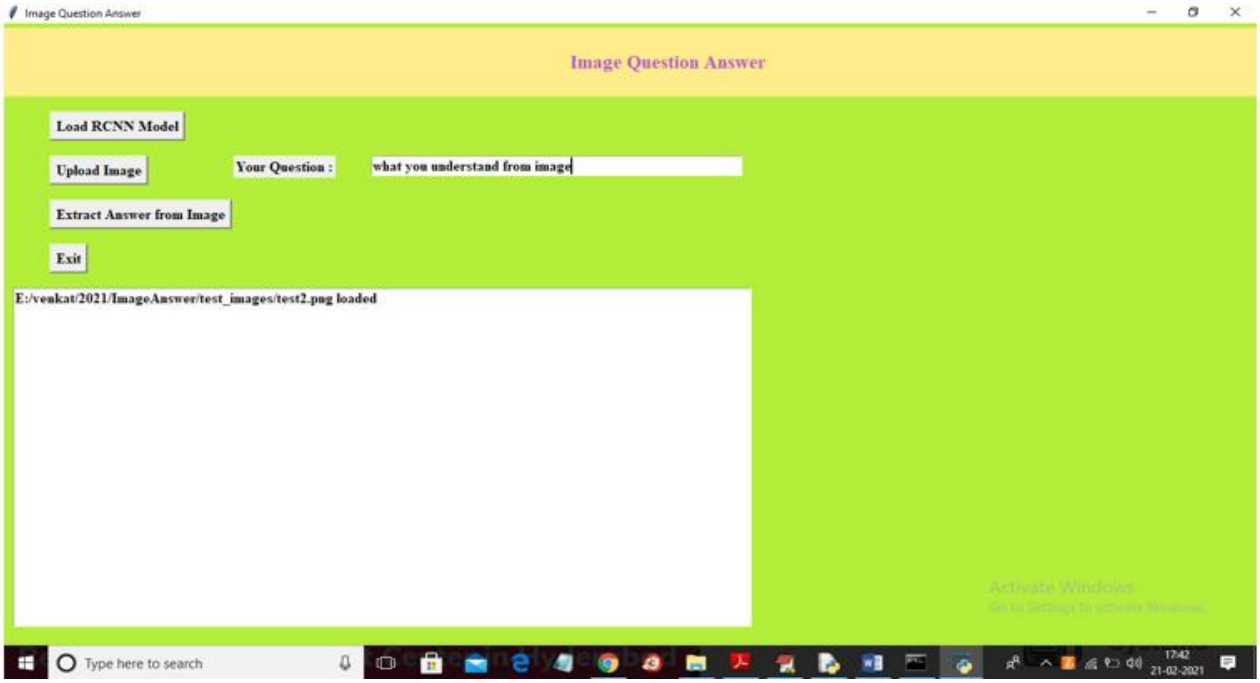


Fig.5In above screen image loaded and now click on ‘Extract Answer from Image’ button to get below result

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

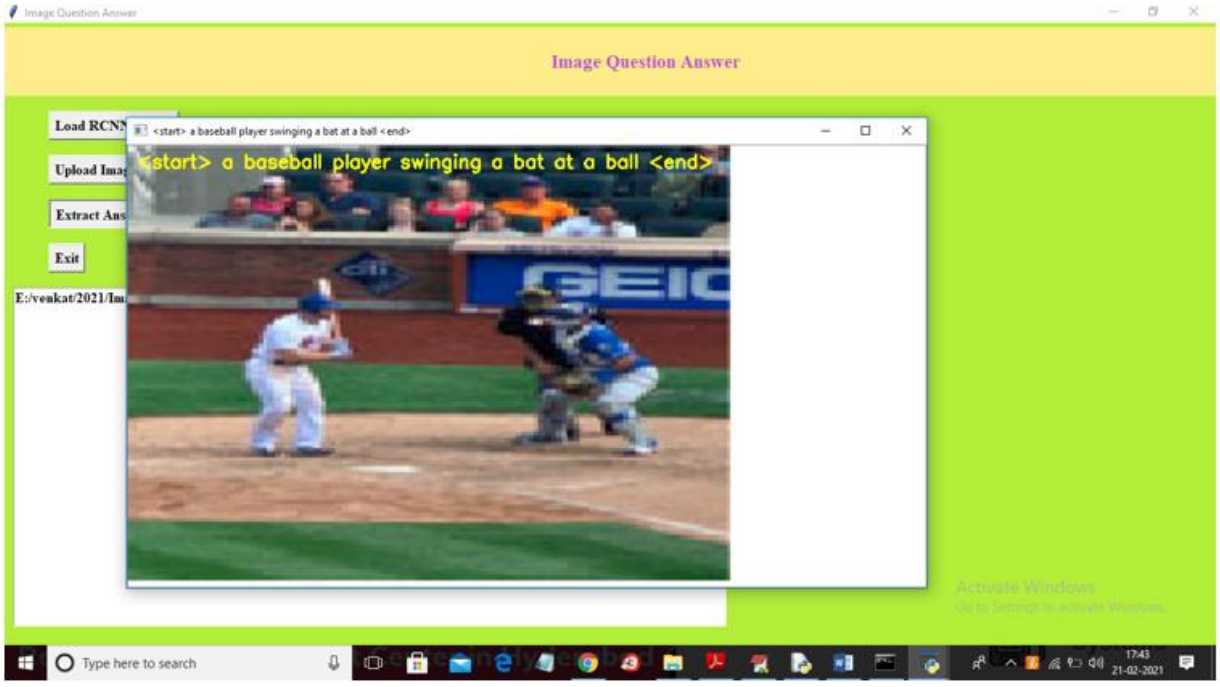


Fig.6 In above screen we can answer from image and similarly you can upload any image and get answer and this project will not give any answer till u entered question

V. CONCLUSION

The VQA problem plays an important part in designing a relatively stringent Visual Turing Test. It requires systems to combine various facets of artificial intelligence like object detection, question understanding and determining the relationship between the objects and the question. Successful Image Question Answering on the whole has been widely known in the community and otherwise to be a momentous breakthrough among artificially intelligent systems. A system that can build a relationship between visual and textual modalities and give matching responses on arbitrary questions, could very strongly mark such a breakthrough. In this paper, we have examined the various datasets and approaches used to tackle the VQA problem. We also examined some of the challenges VQA systems might face in regard to multiword answers and the types of questions asked. Future scope relies on

ISSN: 2057-5688

creation of datasets that rectify bias in existing datasets, introduction of better evaluation metrics for the algorithms carrying out the VQA task to determine whether the algorithm is performing well on the VQA task in general and not only on that dataset. In addition to that, the inclusion of more varied, diverse scenes in the catalogue of images and more realistic, complex questions with a longer, descriptive set of answers would be highly valuable.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

[2] Joseph Redmon, Ali Farhadi "YOLO9000:Better, Faster, Stronger", arXiv:1612.08242v1, 25 Dec 2016.

[3] Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems (NIPS), 2015.

[5] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Lawrence Zitnick, and D. Parikh, "VQA: Visual Question Answering", in Proceedings of the IEEE International Conference on Computer Vision, pp. 24252433, 2015.

[6] M. Malinowski, Mario Fritz, "A multi-world approach to question answering about real-world scenes based on uncertain input ", NIPS, 2014.

[7] N. Silberman, D. Hoiem, et al, "Indoor segmentation and support inference from rgbd images," in

ISSN: 2057-5688

European Conference on Computer Vision (ECCV), 2012.

[8] Kan Chen, Jiang Wang, Liang-Chieh Chen, Haoyuan Gao, Wei Xu, Ram Nevatia, "ABC-CNN: An attention based convolutional neural

network for visual question answering", arXiv preprint arXiv:1511.05960, April 2016

[9] Naga Lakshmi Somu, Prasadu Peddi (2021), An Analysis Of Edge-Cloud Computing Networks For Computation Offloading, Webology (ISSN: 1735-188X), Volume 18, Number 6, pp 7983-7994.