JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

INTELLIGENT EYE USING AI

Ananda Kumar H N*[1], N N Prajwal Kashyap[2], Nithesh Kumar M C[3], Ashitha S[4], Varshini M R[5]
*[1]Assistance Professor, Department of Computer Science & Engineering, Maharaja Institute of Technology Mysore, Karnataka, India
[2,3,4,5]UG Student, Department of Computer Science & Engineering, Maharaja Institute of Technology Mysore, Karnataka, India

Abstract : Experiencing the surrounding, experiencing the nature around for blinds is impossible like that of a normal person. Extracting information from an image or a scene is done by visualizing the picture thoroughly by men, which is a tedious and time consuming task. So we had developed an application which helps the people to get the information of an image in fraction of seconds without even looking at it, using object recognition and voice assistance. our system is so unique that it uses the yolo V3 to detect multiple objects in one shot, which is one of the fastest algorithms to detect the objects. And builds the relationship between the objects detected with the help of RNN, generates a brief paragraph which describes the whole scene in the image using NLP and the audio output is given with the help of pytts package.

***IndexTerms* - Image captioning, yolo V3, RNN, pytts, Automated image captioning, image processing, Automated scene description, NLP**

I. INTRODUCTION

There are many ways depicted in the literature for object detection. Although huge advancement has been made on picture acknowledgment, including both worldwide picture arrangement and neighborhood object discovery, with the help of profound learning strategies and huge scope preparing information, there still exists an immense hole in profound comprehension of pictures.YOLOv3 is the most powerful and advanced algorithm using which object detection can be done with less time. We in this project are aiming to build an application which recognizes the objects (RNN) and builds the relation between the objects recognized and tells it aloud through speakers. It basically builds the relations between the objects detected, and it also describes that scene of the image with a small paragraph. The main target is to make the man independent. It makes the work of the man easy in new places which intern helps in the betterment of the lifestyle. This also makes them self-reliable without asking anyone’s help. This application also helps a lot for blinds to feel the environment better and takes action accordingly.

This can also be customized for any organization and used as guide for new comers or visitors. Where, for example campus tours in institutions and corporate companies. Virtual guides at historical buildings and monuments for tourists using real time object detection. We are achieving this by detecting and recognizing the object then extracting the name of the objects and building the relation between objects, then give the output aloud through speakers.

II. RESEARCH METHODOLOGY

2.1 LITERATURE SURVEY

[1] DEEP STRUCTURED LEARNING FOR VISUAL RELATIONSHIP DETECTION.

This paper was published in the year 2020 by Yaohui Zhu. This paper proposes a significant coordinated model, which acquire a knowledge of relationship with the help of feature level conjecture and name level assumption to additionally foster learning limit of simply making use of feature level predication. The segment level assumption learns relationship by discriminative features, and the imprint level figure learns associations by getting conditions among things and predicates reliant upon the learnt relationship of feature level. Likewise, they use coordinated SVM (SSVM) adversity function as our smoothing out objective, and crumble this goal into the subject, predicate, and article headways which become essential and all the more free. Our tests on the Visual Relationship Detection (VRD) dataset and the colossal extension Visual Genome (VG) dataset endorse the ampleness of our procedure, which outflanks cutting edge strategies.

[2] VISUAL RELATIONSHIP DETECTION WITH LANGUAGE PRIORS

This paper was published in the year 2018 by Cewu Lu**.** This paper, proposes a system that make use of this information to plan visual models for articles and predicates independently and later integrates them to predict various associations per picture. Then upgrade previous work by using language priors from semantic word embeddings to finetune the probability of a normal relationship. Our model can scale to expect countless sorts of associations a few models. In addition, we limit the things in the expected associations as bouncing encases the image. We further show that understanding associations can chip away at content based picture recuperation.

[3] VISUAL GENOME.

This paper was published in the year 2017 by Ranjay Krishna. This paper, present the Visual Genome dataset to empower the demonstrating of such connections. We gather thick explanations of articles, qualities, and connections inside each picture to get familiar with these models. In particular, our dataset contains over 100K pictures where each picture has a normal of 21 articles, 18 ascribes, and 18 pairwise connections between objects. We canonicalize the articles, qualities, connections, and thing phrases in area depictions and questions answer sets to WordNet synsets. Together, these comments address the densest and biggest dataset of picture depictions, objects, characteristics, connections, and question answers.

[4] A HIERARCHIAL APPROACH FOR GENERATING DESCRIPTIVE IMAGE PARAGRAPHS.

This paper was published in the year 2017 by Justin Johnson. In this paper we conquer these impediments by producing whole passages for portraying pictures, which can tell itemized, brought together stories. We foster a model that disintegrates the two pictures and passages into their constituent parts, distinguishing semantic areas in pictures and utilizing a various leveled repetitive neural organization to reason about language. Semantic investigation affirms the intricacy of the passage age task, and intensive analyses on another dataset of picture and section sets show the adequacy of our methodology.

2.2 EXISTING SYSTEM

Despite the fact that critical advancement has been made on picture acknowledgment, including both worldwide picture arrangement and nearby item location with the help of profound learning methods and enormous scope preparing information, we are still a long way from arriving at the objective of far reaching scene understanding, existing models would have the option to recognize discrete articles in a picture however would not be potential to clarify their collaboration or the connections between them. Albeit, these are a few examinations occurred as of late on connections between objects, conditions among articles and predicates have not been completely considered till now. Likewise, in some work just portraying picture with a basic undeniable level sentence, there is a major upper-bound on the quality and amount of data approaches can deliver.

2.3 METHODOLOGY

Our proposed system contains object recognition and relationship forecast. Item identification is utilized to find the locales of articles and a grouping of item matches are orchestrated forecast of relationship. In the relationship forecast, at first we separate the visual highlights of the two items and their association and utilize the intermittent neural organization to produce the depiction.

Item recognition: In the undertaking, we use YOLO V3 to identify the arrangement of articles, and create a grouping of article sets. Then, at that point each item matches with include guide of the picture are acquired for the relationship expectation.

Relationship expectation: The will be the critical part of relationship building are include level forecast and name level expectation. In highlight level forecast it is to acquire include level score of relationship and the mark level expectation is to ascertain the name level score of relationship by catching the conditions between objects. The last score of the anticipated relationship is the loads of amount of the two scores.

Details: To fabricate this framework we considered the fliker8k dataset, an assortment for sentence based picture portrayal and search, comprising of 8000 pictures that are combined with various inscriptions, here pictures are chosen physically to portray an assortment of scenes and circumstances.

We have split those sentences, broken them into individual words and have defined a vocabulary. The image inputed is subjected to yolo algorithm through tensorflow where it uses 106 layers of convolution neural network internally. Then the features are extracted from the images and detects the objects and object pairs in the image. These objects are given to RNN, where objects names are compared with the vocabulary created earlier form the flicker8k model. After the comparision, on different aspects based on the weight score sentences are formed for the input image, describing that image. Finally all the sentences are appended together and displayed as output for the user along with the image. The same is also given as the audio output with the help of PTTS.

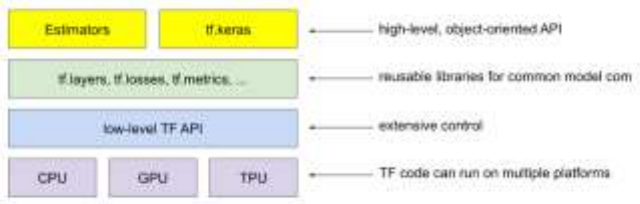
Below are some brief information of some of the packages used in the development of the system:

Tensorflow

TensorFlow is an open-source and free programming library for dataflow and differentiable programming across a degree of endeavors. It is an emblematic numerical library, and is used for artificial intelligence applications like neural affiliations. It can be used for both examination and creation at Google. TensorFlow was made by the Google Brain pack for inside Google use. It was passed on under the Apache License 2.o on November 9, 2o15.

TensorFlow open source platform for artificial intelligence. TensorFlow is a rich construction for working with the whole bits of an artificial intelligence framework; in any case, this class bases on utilizing a specific TensorFlow APIs to make and produce artificial intelligence models. The TensorFlow documentation for complete nuances on the more expansive TensorFlow framework.

TensorFlow API are facilitated consistently, with the basic degree APIs reliant upon the low-level APIs. Man-made knowledge examiners utilize the low-level APIs to make and research new AI calculations. In this class, you will utilize an immense level API named tf.keras to depict and design AI projects and to make suspicions. tf.keras is the TensorFlow assortment of the open-source Keras API. The going with figure shows the request for TensorFlow toolboxs:


Estimators
tf.keras
high-level, object-oriented API
tf.layers, tf.losses, tf.metrics, ...
reusable libraries for common model com
low-level TF API
extensive control
CPU
GPU
TPU
TF code can run on multiple platforms

Pandas

Pandas is an open source Python package that is most usually utilized for information science/information evaluation and AI assignments. It depends on top of another group named Numpy, which offers assistance for multi-dimensional bunches. As maybe the most standard data battling packs, Pandas works outstandingly with various diverse data science modules inside the Python climate, and is typically associated with every Python scattering, from those that go with your functioning system to business trader movements like ActiveState's ActivePython.

Numpy

NumPy is a package used for programming in python language, adding support to gigantic, multi-dimensional showcases and associations, nearby a monstrous assortment of clear level numerical capacities to manage these packs. The earlier version of NumPy, Numeric, was at first made by Jim Hugunin with obligations two or three different designers. In the year 2005, Travis Oliphant made NumPy by joining highlights of the battling Num cluster into Numeric, with wide changes. NumPy is open-source programming and has various providers.

The Python programming language was not at first expected for numerical figuring, notwithstanding pulled in the thought of the consistent and planning neighborhood the get-go. In 1995 the particular vested gathering (SIG) system sig was set up resolved to portray a bunch enrolling pack; amid its people, Python organizer and maintainer Guido van Rossum, extended Python's accentuation (explicitly the requesting etymological construction) to make display handling more straightforward.

Overall Process:

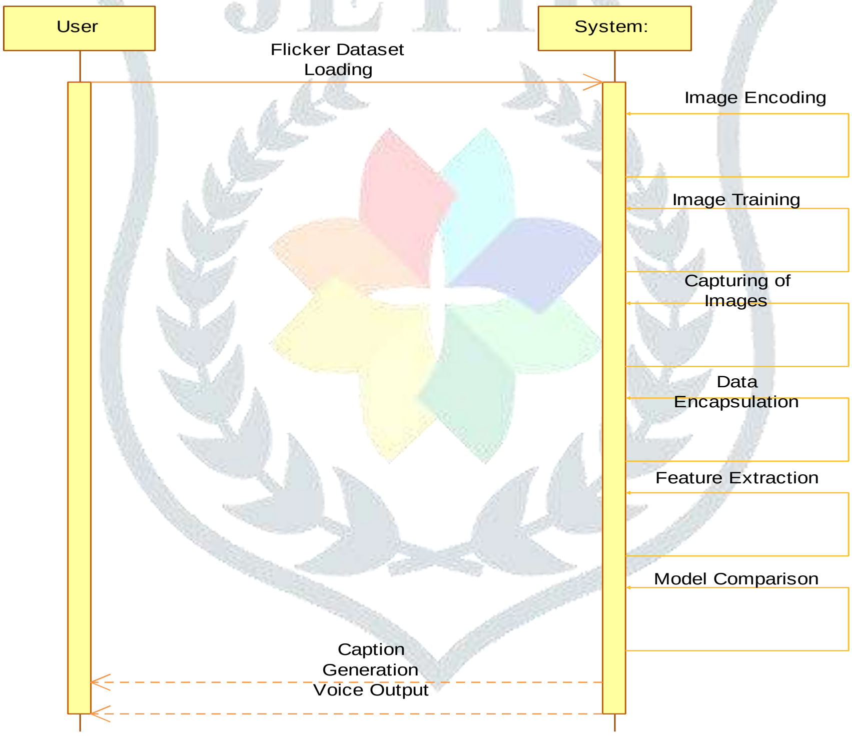


Fig 2: Workflow of the proposed system

The proposed system works as follows:

1. An application will be installed to the computer.
2. User can upload image to that application.
3. User will get audio output which is the description or the caption of that image. Subsequently the caption along with the respective image will be displayed on screen.

2.4 Implementation details

The system is developed in python IDLE 3.9 using image processing technique and Flask. The GUI is developed in flask. The image processing techniques is used for detecting the objects in the given image is YOLOv3, The captions are generated by using RNN(reccurent neural network). The audio output is obtained by using PTTS(Python text to speech) package.

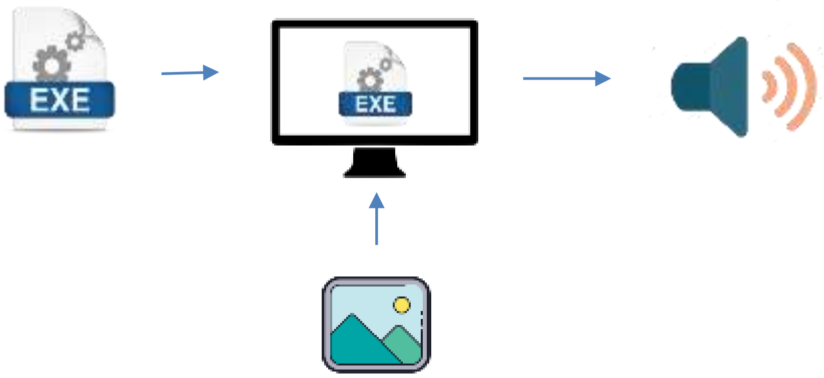


Fig 3: Overall Working of the model

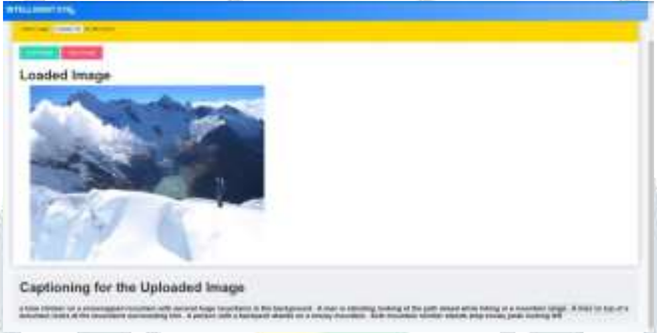
III. RESULTS AND DISCUSSION

3.1 Results

This system gives a complex comprehension of pictures. With our framework, we anticipate that these models should foster a more extensive comprehension of our visual world, supplementing PCs' abilities to recognize objects with capacities to depict those articles and clarify their cooperations and connections.

Our proposed method captures dependencies between objects and builds the relationships among them to generate the paragraph describing the image on the whole, but has not learnt to evaluate the reasonability of detecting relationships.

The results can be obtained with the our simple and easy user interface by clicking on the load button. First the audio output is given, once the audio is complete, the description with the input image will be displayed for the user. The user can clear the image using the clear button to load new image according to the user requirement.



Loaded Image

Captioning for the Uploaded Image

[illegible]

Fig 4: Output

3.2 Conclusion

In the proposed framework, we propose a profound organized model for visual relationship recognition. Our proposed strategy predicts connections on the component level, yet in addition catches conditions among items and predicates. To assess our proposed technique, we direct investigation on Flicker dataset and accomplish cutting edge execution. With utilizing the cover of connections, the test exhibitions are improved. This outlines that assessing sensible connections can work with visual relationship identification and the assessment of sensible relationship is to learn second-order relations of labels.

REFERENCES

[1] Belanger, D., and McCallum, A. 2016. Structured prediction energy networks. In International Conference on Machine Learning, 983–992.

[2] Carreira, J.; Agrawal, P.; Fragkiadaki, K.; and Malik, J. 2016. Human pose estimation with iterative error feedback. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4733–4742.

[3] Chen, L.-C.; Schwing, A.; Yuille, A.; and Urtasun, R. 2015. Learning deep structured models. In Proceedings of the 32nd International Conference on Machine Learning, 1785– 1794.

[4] Dai, J.; Li, Y.; He, K.; and Sun, J. 2016. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, 379–387.

[5] Dai, B.; Zhang, Y.; and Lin, D. 2017. Detecting visual relationships with deep relational networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[6] Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and FeiFei, L. 2009. Imagenet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248–255. IEEE.

[7] Galleguillos, C.; Rabinovich, A.; and Belongie, S. 2008. Object categorization using co-occurrence, location and appearance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1–8. IEEE.

[8] Gkioxari, G.; Girshick, R.; Dollar, P.; and He, K. 2017. De- ´ tecting and recognizing human-object interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[9] Gould, S.; Rodgers, J.; Cohen, D.; Elidan, G.; and Koller, D. 2008. Multi-class segmentation with relative location prior. International Journal of Computer Vision 80(3):300–316.

[10] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778.