Convolutional network based animal recognition using YOLO and darknet.

REDDY, B.K., BANO, S., REDDY, G.G., KOMMINENI, R. and REDDY, P.Y.

2021

Convolutional Network based Animal Recognition using YOLO and Darknet

B. Karthikeya Reddy
Department Of CSE
Koneru Lakshmaiah Education Foundation,Vaddeswaram,India
karthikeyareddy199@gmail.com

Shahana Bano
Department Of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram,India
shahanabano@icloud.com

G. Greeshmanth Reddy
Department Of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram,India
greeshmanthreddy61@gmail.com

Rakesh Kommineni
Department Of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram,India
rakeshkommineni147@gmail.com

P. Yaswanth Reddy
Department Of CSE
Koneru Lakshmaiah Education Foundation,Vaddeswaram,India
pallapoluyaswanthreddy@gmail.com

Abstract— In general, the manual detection of animals with their names is a very tedious task. To overcome this challenge, this research work has developed a YOLOV3 model to identify the animal present in the image given by user. The algorithm used in YOLOV3 model is darknet, which has a pre-trained dataset. The overall performance of the model is based on different training images and testing images of the dataset. The main goal of this research work is to build an animal recognition methodology using YOLOV3 model. The image of animal will be given as input, then it will display the name of the animal as output by using YOLOV3 model. The detection is done by using a pre-trained coco dataset from darknet.

Keywords— YOLO V3, Darknet, Convolutional network, Detector, Opencv

I.INTRODUCTION

YOLO (You Only Look Once) it is an object detector which uses the feature of deep convolutional neural network [1]. YOLO is a fully convolutional network. YOLO consists of 75 convolutional layers with skip connections [2]. When an input image has been considered, it goes through the feature extraction so it is divided into different scales. Its output is generated by applying a 1x1 kernel on feature map. These features are going into detector to bounding boxes information. Machine learning is a sub-branch of artificial intelligence, it has been applied to image processing [3]. It helps us to implement object detection to detect and recognize the objects from our given input images and video [4].

The feature extractor for YOLO V3 model is Darknet-53. In the previous version of YOLO model there're only 19 layers so the network extends the layers from 19 to 53 layers for YOLO V3 model[5]. For detection of input image 53 more layers are added to the YOLO V3 model [6]. YOLO V3 is designed to be a multi-scaled detector.

Once we have features, we can add them into detector. Through 1x1 and 3x3 Convolutional layers are used to form the final output for different scales[7]. It also consider the features of previous scales by doing small scale detection and large scale detection[6]. The feature of YOLO V3 is to make detection at different scales[7]. YOLO V3 model uses total of 9 anchor boxes i.e; each box consists of three different scales[8]. For each scale it can predict 5 boxes using 3 scales[9]. Convolutional Neural Networks (CNN) Image Processing Algorithms which has been properly trained to identify the animal image[10].

Darknet-53 is a convolutional neural network that is of 53 layers [11]. You can load a pre-trained version of the network which is trained on different images [12]. The pre-trained version classifies images into object categories [13]. Darknet-53 is the foundation of Object detection and YOLO workflow [14]. Darknet-53 will give us 53 more layers has been added total of 106 layers are been consider for Darknet-53[15]. This process overcomes the slowness of the previous version of the model [16]. Input image is recognized using features on the face of animal [17].

II. RELATED WORK

There are some papers which are using YOLO model with keras. In the paper which we have studied we noticed that it consumes more time to predict the output and we cannot select our own images. So, by using our model we can implement our own no of images of our own choice.

YOLO makes uses only convolutional layers, making it a totally convolutional network. YOLO is Build with 75 convolutional layers, with considering skip connections and upsampling layers. For example, if the network is having 32 layers, then an input image of size 416 x 416 will yield an output of size 13 x 13[10].

Darknet is an open source framework written in C and CUDA. It is fast, easy to put in, and supports CPU and GPU computation. Darknet is an open source framework written in C and CUDA. It is fast, easy to put in, and supports CPU and GPU computation.YOLOv3 predicts an objectness score for every bounding box using logistic regression [15]. Each ground truth object is related to one boundary boxes.

III. PROPOSED WORK

Importing all the packages:

To make our model run we need to install some of deep learning packages. These packages has numerous amounts of inbuilt hidden classes. One among those classes is recognizer which makes them the system to recognize what the source is trying to know. It is trained using various APIs that are available in these packages. There are many APIs in the recognizer class, some of them can be used online and one among them is used when the computer system is working in offline mode. Here in our model, we have used google API which recognize the images that the user selected as input.

Later we need to install the OpenCV package where it deals with detection class. That will reads the input image and stores the image details. When we need the image it will uses the help of packages that we installed.

For the whole purpose of creating output and the visualization of the entire model, some of the packages need to be installed like matplotlib. This makes the output to be more effective which displays the desired image with text.

Working Principle:

Using the given input image by the user the system will displays the output with text and accuracy. Later the image is broken into various lengths and widths based on the given input image. Here for the recognition of image, YOLOV3 model is using recognizer deep learning package. The recognizer class enables the model to recognize with the help of an online tool that is google API which is used for recognizing different images. This model displays the name of animal for the output image. To display the output image we used the matplotlib package which plot a box around the animal.

Pseudo code:

Step-1: Start
Step-2: Reading the image using OpenCV package.
Step-3: If the model received a image then the model goes to the recognizing phase.
Or Else asks the user to give image as an input.
Step-4: Darknet comes into the functionality.
Step-5: Recognition of image with name of animal in the desired language for user.
Step-6: Displays the output image.
Step-7: Repeat the steps 1,2,3,4,5,6.
Step-8: End

IV.FLOW CHART

Firstly, the darknet is cloned into repository and then enables the GPU and OPENCV, where GPU is used for running the model constantly and then it moves to CUDA Compiler Driver which helps to hide the integrated work and after that it builds the darknet and it downloads the pretrained dataset. Then we will read the image and then it uploads the files to darknet and automatically it downloads the files from darknet then it mounts the google drive and predicts the image with name and accuracy.

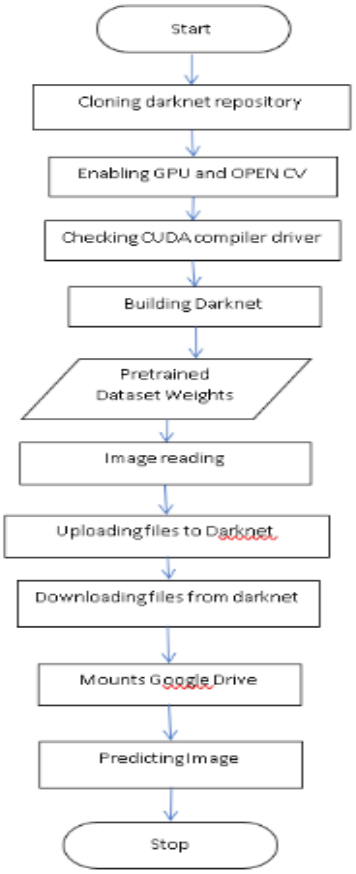

Fig-1: Flow Chart

V. BLOCK DIAGRAM

First we are reading the images from the user then the images has been sent to feature extractor to divide the image into different scales these scales has been connected to the detector phase in this the images are been recognized and then the bounding boxes will come to functionality. In this feature extractor is used for image processing and for detect images or video stream..

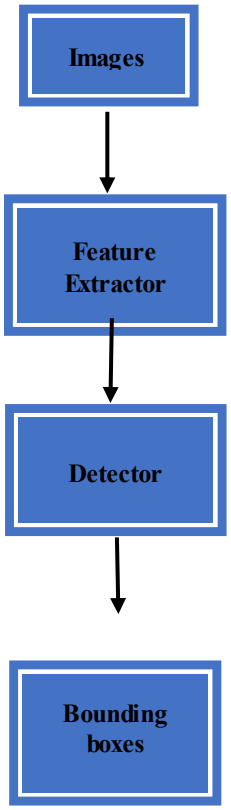

Fig-2: Block Diagram

VI.RESULT ANALYSIS

By using this model we have recognized the animal through image and also comparison is done on types of animals table as follows.

Input image	Status	Output image	Accuracy	Performance (in milli seconds)
Bear	Correct Output	Bear	1.00	18.402000
Cat	Correct Output	Cat	0.99	18.483000
Crow	Correct Output	Bird	0.94	14.720000
Dog	Correct Output	Dog	0.87	18.424000
Eagle	Correct Output	Bird	1.00	18.288000
Giraffe	Correct Output	Giraffe	1.00	18.362000
Hen	Correct Output	Brid	0.56	18.432000
Ostrich	Correct Output	Bird	1.00	18.384000
Horse	Correct Output	Horse	1.00	18.354000
Zebra	Correct Output	Zebra	1.00	18.314000
Penguin	Correct Output	Bird	0.50	18.367000
Wolf	Correct Output	Dog	0.99	18.115000
Yak	Correct Output	Cow	0.96	18.395000
Peacock	Correct Output	Bird	0.98	18.362000
Elephant	Correct Output	Elephant	0.40	18.667000
Penguin	Correct Output	Bird	0.98	18.459000
Elephant	Correct Output	Elephant	0.91	18.455000
Sheep	Correct Output	Sheep	1.00	18.359000
Cow	Correct Output	Cow	0.46	20.445000
Bear	Wrong Output	Dog	0.77	18.401000
Alligator	Wrong Output	Bird	0.27	18.329000
Chimpanzee	Wrong	Bird	0.48	18.434000

	Output			
Fox	Wrong Output	Cat , Dog	0.32,0.61	18.406000
Fox	Wrong Output	Dog , Horse	0.91,0.57	18.323000
Hippopotamus	Wrong Output	Cow	0.47	18.361000
Jaguar	Wrong Output	Bird	0.67	18.490000
Kangaroo	Wrong Output	Dog	0.96	18.411000
Lion	Wrong Output	Dog	0.99	18.399000
Monkey	Wrong Output	Dog	0.95	18.368000
Rabbit	Wrong Output	Cat , Bird	0.48,0.35	18.430000
Spider	Wrong Output	Bird	0.35	18.284000
Tiger	Wrong Output	Zebra	0.97	18.488000
Tiger	Wrong Output	Zebra , Giraffe	0.42,0.81	18.431000
Rabbit	Wrong Output	Cat	0.57	18.503000
Chimpanzee	Wrong Output	Dog	0.99	18.271000
Fox	Wrong Output	Dog	0.98	18.398000
Hippopotamus	Wrong Outpu	Elephant	0.95	18.361000
	t			
Kangaroo	Wrong Output	Cat , Dog	0.34,0.96	18.292000
Monkey	Wrong Output	Cat , Bird	0.41,0.45	18.427000
Spider	Wrong Output	Potted Plant	0.72	18.288000
Tiger	Wrong Output	Dog	0.76	18.348000
Donkey	Wrong Output	Horse , Cow	0.96,0.59	18.404000
Impala	Wrong Output	Cow	0.74	18.530000
Fish	Wrong Output	Bird	0.96	18.605000
Lion	Wrong Output	Dog	0.59	18.293000
Alligator	Wrong Output	Bird	0.33	18.426000
Lion	Wrong Output	Cow , Dog	0.38,0.69	18.279000
Deer	Wrong Output	Sheep	0.25	18.375000
Sheep	Wrong Output	Sheep , Cow	0.97,0.32	18.403000
Goat	Wrong Output	Cow , Giraffe	0.45,0.93	18.492000
Impala	No Output			18.362000
Jaguar	No Output			18.411000
Scorpion	No Outpu			18.516000

	t			
Snake	No Output			18.564000

Here, Right output means the image which are predicted correct type of animal name were as Wrong output means the images which are predicted a different name rather than the correct name of the given input image. No output means it is not able to predict the given input images. Some of the outputs with right prediction, wrong prediciton and also no output images are as follows.

No Predictions:

By giving the following fig-4 the model has not displayed any output of the animal image.

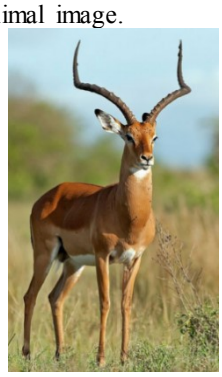

Fig-4 : Deer image

By giving the following fig-5 the model has not displayed any output of the animal image.

Fig-5 : Scorpion image

By giving the following fig-6 the model has not displayed any output of the animal image.

Fig-6 : Snake image

By giving the following fig-7 the model has not displayed any output of the animal image.

Fig-7 : Jaguar image

VII.CONCLUSION

By training this model we have learned to how the YOLO V3 model works and identifies the animal. When we run the YOLO V3 model then it is predicting right predictions and wrong predictions and for some images it is not predicting and also it is not taking input images correctly. So to overcome this problem in Future, we can train the model using custom dataset for our own no of classes and no of images in each class to get better accuracy.

VIII.FUTURE SCOPE

We can train the model using custom dataset for own no of classes and no of images in each class to get better results.

REFERENCES

[1] Michal Maj "Object detection and Image Classification with YOLO" KDnuggets , September 2018.

[2] Ayoosh Kathuria "What's new in YOLO v3?. A review of the YOLO v3 Object" Towards data science , April 23rd 2018.

[3] Ethan Yanjia Li "Dive Really Deep into YOLO v3: A Beginner's Guide" Towards data science, December 31st 2019.

[4] Joseph Redmon , Ali Farhadi "YOLO v3: An Incremental Improvement" University of Washington, April 8th 2018.

[5] Manogna Mantripradaga "Digging deep into YOLO v3 – A hands-on guide" Towards data science, August 16th.

[6] Joseph Redmon "Darknet-53 Explained" Cornell University, April 2018, https://paperswithcode.com/method/darknet-53.

[7] Ayoosh Kathuria "How to implement a YOLO v3 Object detector from Scratch in PyTorch" KDnuggets, May 2018.

[8] S. Sandeep Kumar, Shruti Mishra, R. Sai Sandeep, U. Sai Ravi Teja, Pradeep Kumar M, M. Shruti, K. Shravya "Deep learning based image recognition for vehicle number information" International Journal of Innovative Technology and Exploring Engineering, Volume: 08, Issue: 8S2, June 2019.

[9] R. Abinaya, Lakshmana Phaneendra Maguluri, S. Narayana, Maganti Syamala "A Novel Biometric Approach For Facial Image Recognition using Deep Learning Techniques" International Journal of Advanced Research in Engineering and Technology (IJARET), Volume: 11, Issue: 9, September 2020.

[10] P. Meghana, S. Sagar Imambi, P. Sivateja, K. Sairam "Image Recognition for Automatic Number Plate Surveillance" International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume: 08, Issue: 04, February 2019.

[11] Pabbisetty Nikhitha, Palla Mohana Sarvani, Kanikacherla Lakshmi Gayathri, Dhanush Parasa, G Yedukondalu "Detection of Tomatoes Using Artificial Intelligence Implementing Haar Cascade Technique" International Conference on Communication, Computing and Electronics Systems, 2020 – Springer.

[12] Sai Manvitha Chittajallu, Navya Lakshmi Deepthi Mandalaneni, Dhanush Parasa, Shahana Bano "Classification of Binary Fracture Using CNN" Global Conference for Advancement in Technology (GCAT), October 2019.

[13] Shahana Bano Vishal P., Snigdha L.K. "An Efficient Face Recognition System using Local Binary Pattern" International Journal of Recent Technology and Engineering (IJRTE), February

2019.
[14] K Bhagavan, S Sagar Imambi, Shahana Bano "A compressive survey on different image processing techniques to identify the brain tumor" International Journal of Engineering & Technology, Volume: 07, Issue: 2.7, 2018.
[15] "Machine-learning system tackles speech and object recognition" https://news.mit.edu/machine-learning-image-object-recognition-0918.
[16] Manoharan, Samuel. "Embedded Imaging System Based Behaviour Analysis of Dairy Cow." Journal of Electronics 2, no. 02 (2020): 148-154.
[17] Vijayakumar, T., and R. Vinothkanna. "Capsule Network on Font Style Classification." Journal of Artificial Intelligence 2, no. 02 (2020): 64-76..