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

Object Detection Using YOLO Technique

[1]Sarah Muskan, [2]Dr. Jalesh Kumar
[1]Student, [2]Professor
[1]Dept. of Computer Science and Engineering, JNNCE
[2]Dept. of Computer Science and Engineering, JNNCE, Shivamogga, India

Abstract : **The object detection task entails locating and identifying certain objects on a picture and classifies them. It is the capacity of software systems to identify objects in an image or a video. Various technologies use object detection such as detection of faces and vehicles, security systems, driverless cars etc. Real-time processing is designed using an object detection system technology called YOLO(You Only Look Once). It has several advantages when compared to that of other object detection methods. Various algorithms like Convolutional Neural Network and Fast Convolutional Neural Network, do not look at the image completely, but Yolo does. In this dissertation, it is proposed that the model predicts the objects by drawing bounding boxes around them using convolution networks and detects the objects faster compared to other algorithms.**

***IndexTerms* – YOLO, Image, Video, Real-Time.**

I. Introduction

Until a few years ago, the development of user interfaces for software and hardware image processing systems accounted for the bulk of programmers employed by each company. Since the debut of the Windows operating system, a lot has changed, and the bulk of developers now focus on finding fixes for image processing problems. The ability to solve common problems, such as recognition of faces, numbering of vehicles, reading of traffic signs, telemetry, and medical image interpretation has nonetheless made substantial strides. Each of these age-old issues was finally resolved after much trial and error involving numerous teams of engineers and scientists[1].

The job of automating the creation of software tools to tackle intellectual difficulties has been strongly envisioned and solved due to the high cost of contemporary technological solutions. To enable the study and identification of previously unidentified objects in images as well as the effective development of applications by regular developers, the field of image processing requires a specific set of tools. For instance, Windows Toolbox enables the creation of interfaces to address numerous real-world issues[6][13].

Object Detection is a branch of computer science. Image processing and computer vision covers two well-researched subfields of object detection such as face detection and pedestrian detection. Many computer vision applications, including image retrieval and video monitoring, can benefit from object detection.

Computer vision is an interdisciplinary field of study that examines how effectively computers can comprehend digital images or movies. The field of technology known as computer vision uses its hypothesis and models to create computer vision systems. Computer vision includes the subfields of reconstruction of scene, detection of events, tracking of videos, object recognition, learning, indexing, motion estimation, image serving, 3D scene modeling, and picture restoration[2].

The "you only look once" (or "YOLO") algorithm is well known for its great accuracy when used in real-time. It outputs recognised items along with bounding boxes after non-max suppression. When it comes to bounding box localization and classification, YOLO uses a single CNN network. You Only Get One Look: Real-time, unified object detection.

There are many object detection techniques, however the Yolo methodology has a number of advantages over other object detection strategies. Other algorithms, including Convolutional Neural Network and Fast Convolutional Neural Network, do not look at the image completely but YOLO recognizes the image faster than other algorithms because it predicts the bounding boxes using convolutional networks with higher accuracy[7][8].

II. Design and implementation

2.1 YOLOv5 Architecture

The YOLO v5 architecture consists of 3 parts namely: 1. Backbone 2. Neck 3. Head.

The below figure 1 shows the architecture of YOLOv5.

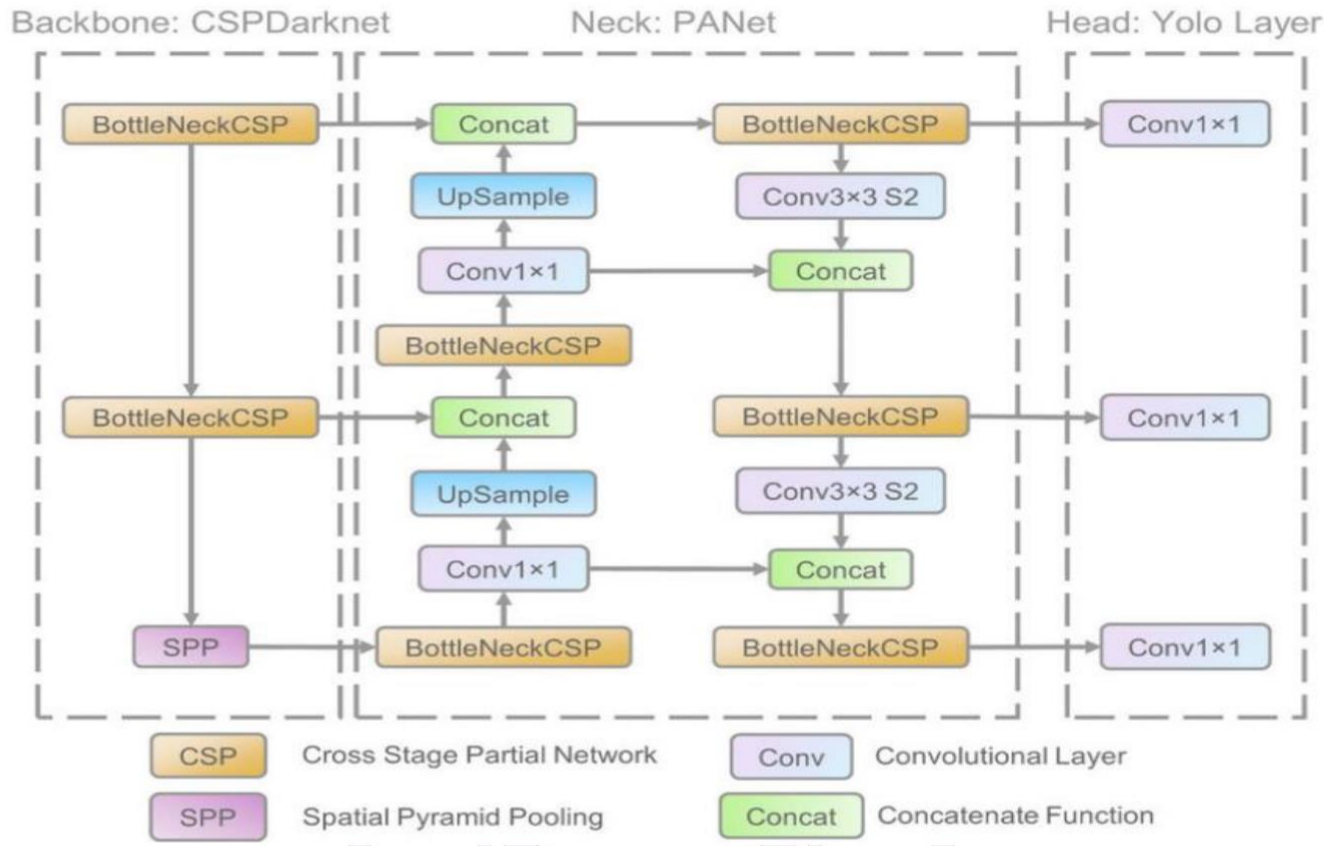

Figure 1: YOLO v5 Architecture

Neck: For the creation the signature pyramids this part of the model is used. With regard to object proportions, feature pyramids aid in the successful generalisation of models. The definition of the same thing in many scales and sizes is helpful. Pyramids of features are a great tool for making models operate effectively with fresh data[3].

Head: It is incharge of detecting things in final stage. It constructs final output vectors using anchor boxes containing class probabilities, objective points, and bounding boxes[12].

Cross Stage partial Network: It is a backbone that increases the learning capacity of CNN.With the help of the Cross Stage Partial Network (CSPNet), the complexity of network optimization can be significantly decreased while preserving accuracy by attributing the issue to redundant gradient information.

Convolutional Layer: The foundation of a CNN is a convolutional layer. It features several filters (or kernels), the parameters of which must be learned during training. In most cases, the size of the filters is smaller than the original image. After each filter convolves with the picture, an activation map is produced. The dot product between each filter element and the input is calculated at each spatial point as the filter is moved over the image's height and width for convolution. [4][9].

Spatial Pyramid Pooling: It is a pooling layer. Constraints that are fixed sized are removed in the network. The fully-connected layers receive the fixed-length outputs produced by the SPP layer, which pools the features (or other classifiers).

2.2 Dataset

For training purposes, the COCO Dataset is used. The COCO dataset consists of three subsets: range, segmentation, and subtitle. The object segment is where the analysis of image borders and the items contained inside them begins. When tagging an image, boundary boxes are employed. Underlying correlation architecture between the image and the things it contains is represented after in-context recognition. Following this, the same colour or grayscale is combined. Here is a section of the super pixels where they can be used to minimize entry items for calculations as well as to help emphasize significant locations[5]. The COCO dataset consists of 5 parts those are:

1. Info: It consists of general information about the dataset.
2. Licenses: It consists of license information.
3. Images: It consists of list of images present.
4. Annotations: It consists of various annotations that are present in the images. These annotations also includes bounding boxes around the picture.
5. Categories: It consists of various categories. This can be used for labeling purpose.

2.3 Implementation

This model detects the object faster compared to that of other algorithms and this can be achieved by following steps:

1. Giving the image/video as the input.
2. Utilizing convolutional networks and the class probabilities for these boxes to predict the bounding boxes.
3. Detecting the objects.

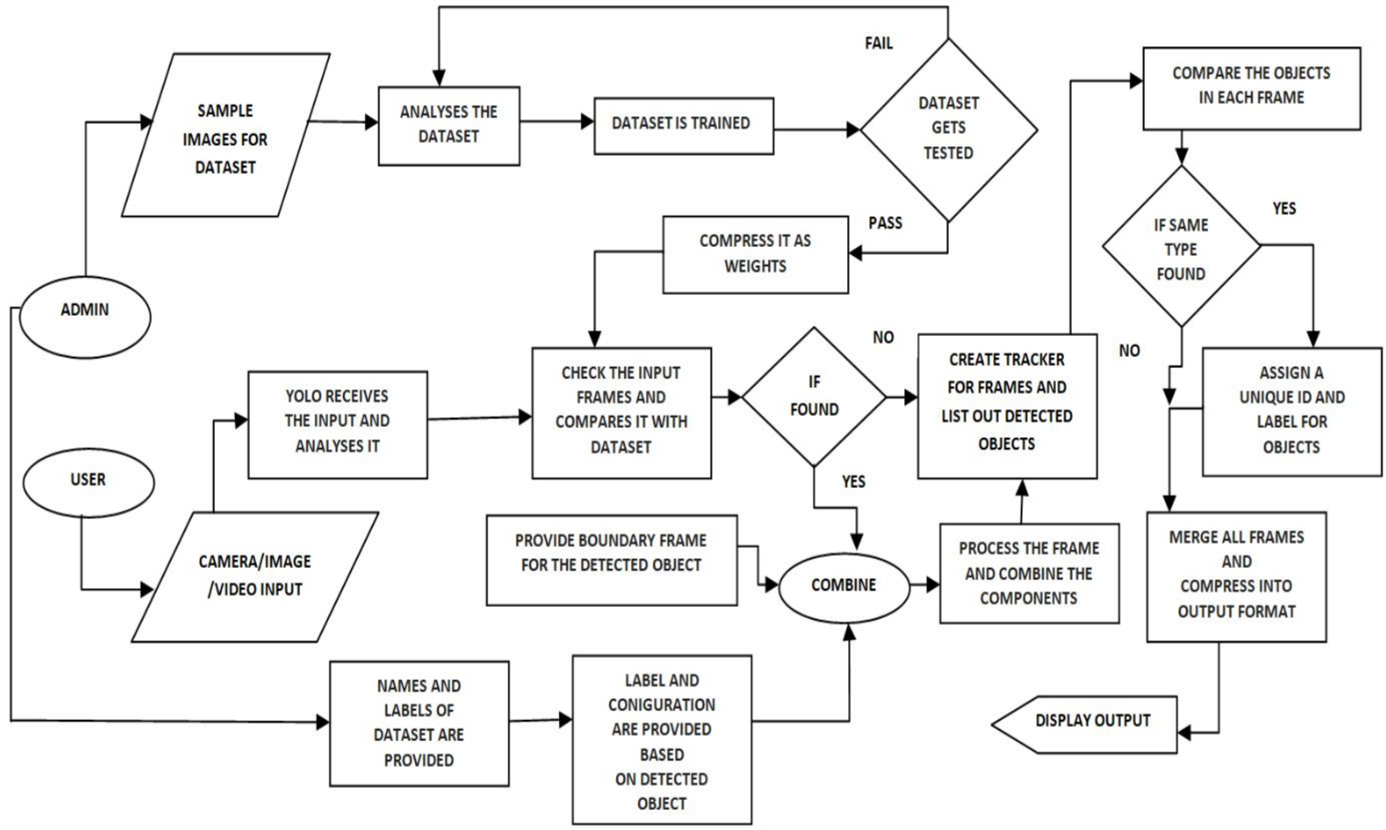

Figure 2: Implementation flow chart

The model or the administrator is shown with sample photos from the dataset in figure 2 above. If the tested dataset is accurate, the dataset is tested later and compared as weights. Data is divided into frames using input from a camera, video, or image, and then each frame is delivered to the Yolo detection algorithm with a user-selected model [5].

A custom model for detection can be made or a predefined model, such as the COCO dataset, can be used. Once an object has been located, the detection process is delineated by boxes and labels before being delivered to the output, where all of the identified frames are gathered and then compressed into the output size[10][11].

III. RESULTS AND ANALYSIS

3.1 Results

The system starts by opening the camera to capture a video or feed the recorded video or image as input to the system of object detection. The imported data is then processed in frame for detection. Yolov5 uses dataset to detect the objects that are present in the input data during detection process. Once the objects are detected they are classified and represented using bounding boxes and labels around the objects and then it is converted into the output.

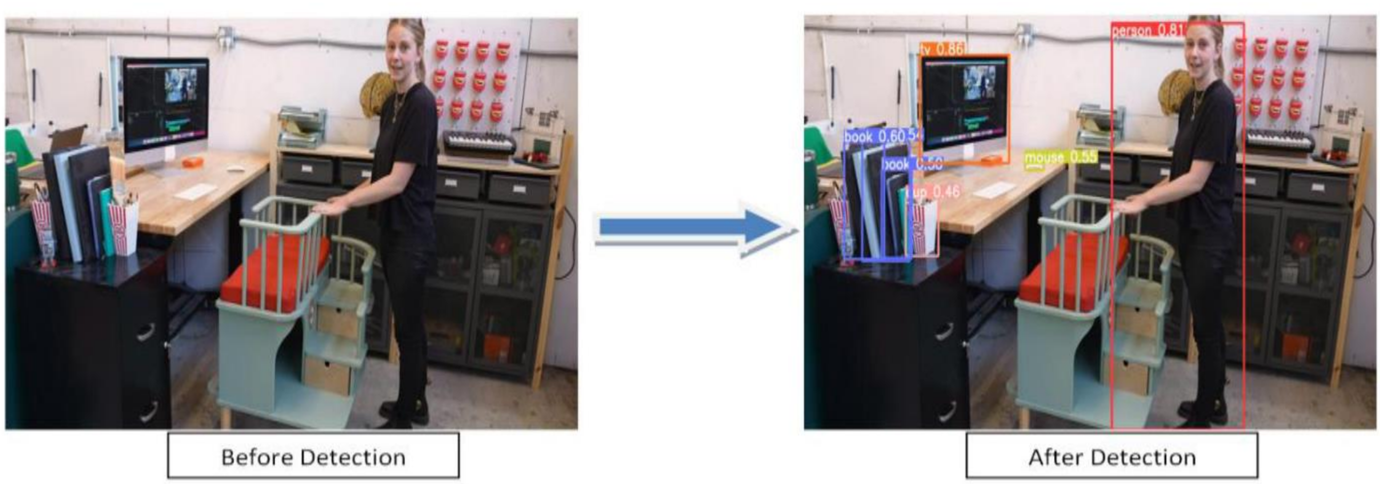

Figure 3: Detection of objects in an image

In the above figure 3, the models recognizes a person, a TV, 2 books, a cup and a mouse with an accuracy of 81%, 86%, 60%, 50%,46% and 55%.

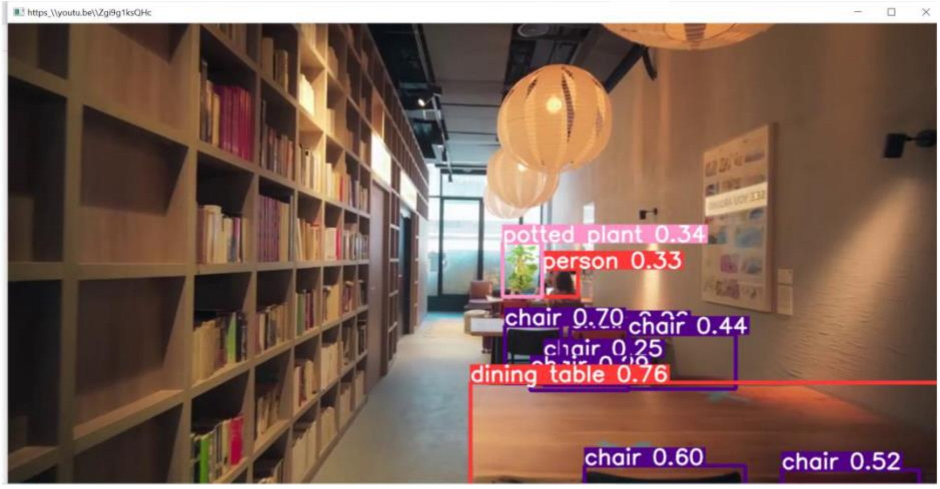

Figure 4: Detection of objects in a video

In the above figure 4, the model recognizes 6 chairs, a dining table, a person and a potted plant with 70%, 44%, 60%, 55%, 25%, 29%, 76%, 33% and 34%.

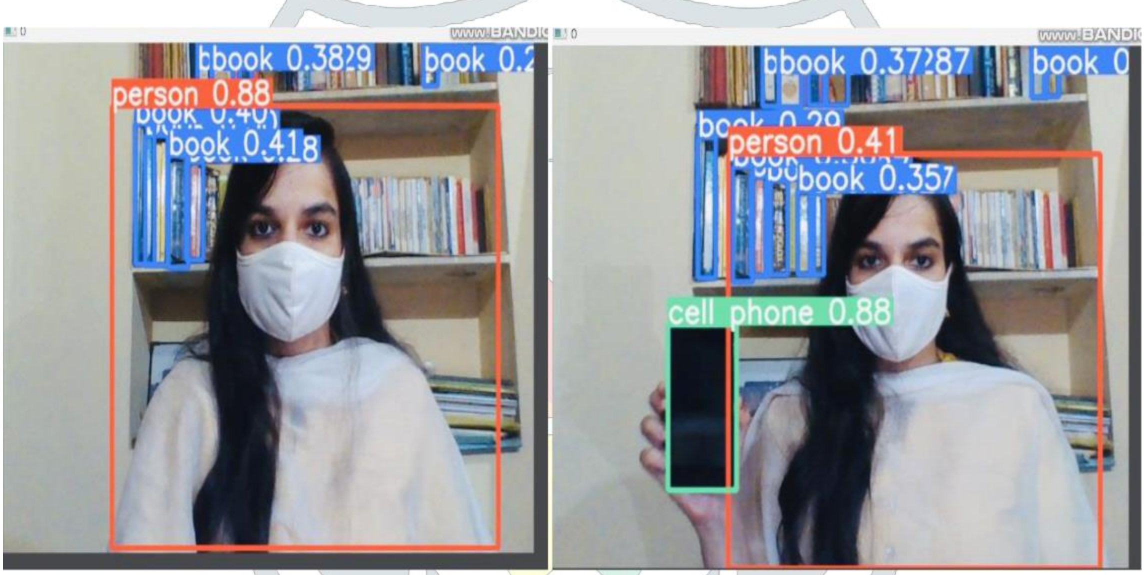

Figure 5: Webcam-based real-time object detection

In the above Figure 5, the model recognizes the real-time objects in the webcam and displays the output on the screen. It recognizes a person and some books with an accuracy of 88% and 60%. It also recognizes a cell phone, a person and some books with an accuracy of 88%, 45% and 60%.

3.2 Analysis

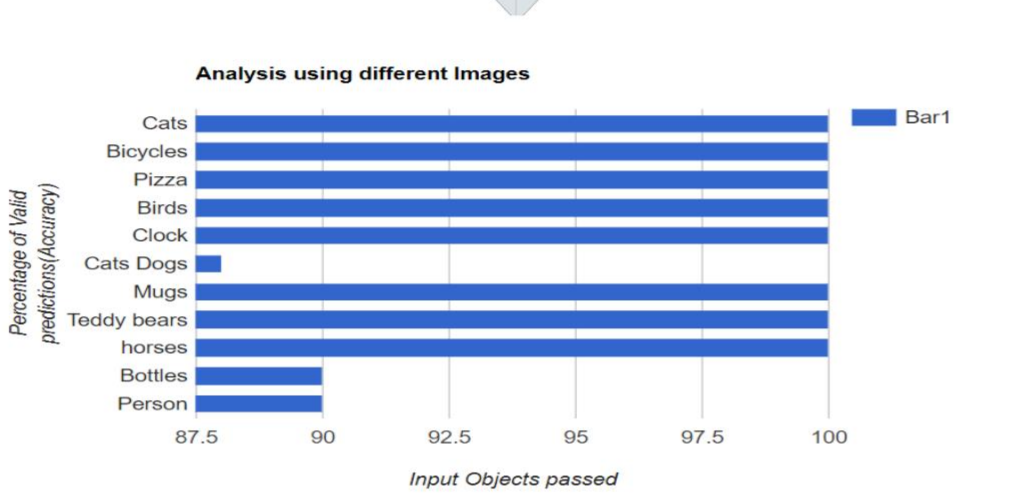

Figure 6: Graph plotted based on the Analysis of different Images

In the above Figure 6, a graph is plotted based on analysis of different images. The model recognizes images of cats, bicycles, pizza, birds, clock, mugs, teddy bear, horses with 100% accuracy. It detects images of cats with dogs with 88% accuracy. It also detects images of bottles and person with 90% accuracy.

IV. CONCLUSION

The YOLO technique is suggested in this research with the goal of object detection using a single neural network. This method is general and performs better than other approaches when used to natural photos in different disciplines. The algorithm may be trained on an entire image and is straightforward to build. Limits are anticipated by YOLO by using the complete image. It also forecasts a decrease in background false positives. Comparing this approach to existing classification algorithms, it is noticeably more effective and quicker to use in real time. Similar to the first Industrial Revolution, object detection technology's promise to relieve people from mundane tasks that machines can perform more effectively and efficiently is currently being tested. Additionally, it will open up fresh directions for investigation and work that will result in future gains.

REFERENCES

[1] Hammad Naeem, Jawad Ahmad, Muhammad Tayyab, "Real-Time Object Detection and Tracking", IEEE Paper, 2013

[2] Sudharshan Duth P, Swathi Raj,"Object Recognition in Images using Convolutional Neural Network ", Proceedings of the Second International Conference on Inventive Systems and Control IEEE, 2018

[3] Mingyuan Xin, Yong Wang, "Research on image classification model based on deep convolution neural network", Xin and Wang EURASIP Journal on Image and Video Processing Springer, 2019

[4] Ivan Fomin, Vladislav Burin, Aleksandr Bakhshiev, "Research on Neural Networks Integration for Object Classification in Video Analysis Systems", International Conference on Industrial Engineering, Applications and Manufacturing IEEE, 2020

[5] Hicham RIRI, Abdelmajid ELMOUTAOUAKKIL, Abderrahim BENI-HSSANE, Farid BOURZGUI, "Classification and recognition of dental images using a decisional tree", 13th International Conference Computer Graphics, Imaging and Visualization, IEEE, 2016

[6] William Tarimo, Moustafa M.Sabra, Shonan Hendre, "Real-Time Deep LearningBased Object Detection Framework", IEEE Symposium Series on Computational Intelligence, 2020

[7] Kushal M, Kushal Kumar B V, Charan Kumar M J, Prof.Pappa M, "ID Card Detection with Facial Recognition using Tensorflow and OpenCV", IEEE Proceedings of the Second International Conference on Inventive Research in Computing Applications, 2020

[8] Mansi Mahendru, Sanjay Kumar Dubey, "Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3", IEEE 11th International Conference on Cloud Computing, Data Science & Engineering, 2021

[9] Ashwani Kumar, Zuopeng Justin Zhang, Hongbo Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm", Sprinhger Open Kumar et al. EURASIP Journal on Wireless Communications and Networking, 2020

[10] A.A.M. Al-Saffar, H. Tao, M.A. Talab, Review of deep convolution neural network in image classification (International conference on radar, antenna, microwave, electronics, and telecommunications). IEEE, Jakarta, 2018 42

[11] Upesh Nepal, Hossein Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs", Sensors, 2022

[12] Chien-Yao Wang, Hong-Yuan Mark Liao1, Yueh-Hua Wu1, Ping-Yang Chen, Jun-Wei Hsieh, I-Hau Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN", IEEE,CVPR Workshop, 2020

[13] Geethapriya. S, N, Duraimurugan, S.P. Chokkalingam, "Real Time Object Detection with Yolo", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8, 2019