International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 123-129
Published Online June 2023 in IJEAST (http://www.ijeast.com)

IDENTIFICATION OF MONUMENTS FROM AERIAL IMAGES USING DEEP LEARNING TECHNIQUES

Gautam Kumar Jaiswal
Student, HMRITM, Delhi, India

Renu Chaudhary
Assistant Professor, HMRITM, Delhi, India

Mohit
Student, HMRITM, Delhi, India

Srishty
Student, HMRITM, Delhi, India

Abstract: **In the continuous development of object detection technology, various deep learning models with high accuracy and speed have been used to detect the various things such as cars, buildings, planes, etc. The aims of this paper to establish a real-time monuments identification from UAVs that can help archaeological surveys and researchers. Project focuses on the implementation of popular object detection algorithms VGG-16 and YOLOv5 and compares the speed, accuracy, and performance of both algorithms, and finds the best fit for real-time detection from an aerial view. Ultimately, increased the surveys with low cost, generation of large accurate UAV data in less time followed by better quality and supply that meets the researcher's demand.**

Keywords: **ATEET, Deep Learning, Aerial Imagery, UAV, Object Detection, Monuments, VGG-16, YOLOv5.**

I. INTRODUCTION

Delhi, The capital of India and the center of heritage, individual architectural monuments and historic urban centers. Monitoring and developing the application to identify and locate these sites from an aerial view. Drones and autonomous planes are used to acquire aerial photographs and collection of high-resolution imaging data in a stipulated time at a low cost and revolutionize the practice by allowing surveyors to detect possible areas of interest, before going messy. With the combination of aerial data and deep learning, identification and localization are quite easy which is helpful to connect with other technologies such as IoT, AR, and VR.

The objective of the research focused to prepare the custom aerial imagery dataset of some monuments of Delhi (Qutub Minar, Lotus Temple, Jama Masjid, Humayun's Tomb) and implementing deep learning techniques to detect the monuments. VGG16 and YOLO v5 model is used for monument detection and compared the accuracy to choose the best analysis model. We can develop a simple user interface of the web application to demonstrate our research project in which users can easily upload an image and get the result of detection with accuracy.

1.1 Unmanned Aerial Vehicle (UAV)

The acronym of UAV is Unmanned Aerial Vehicle, which is commonly known as a drone. It is an aircraft without a human pilot on board, controlled remotely by a person on the ground or autonomously by a pre-programmed system which have embedded systems, working with onboard sensors and Global Navigation Satellite Systems (GNSS) [1] as an example shown in Fig.1. UAVs are playing important part in number of applications including but not limited to surveying and mapping, precision agriculture, monitoring and inspection, advertising, emergency, archaeology, response and disaster management, defense, urban planning, healthcare, waste management, weather forecasting, mining and telecommunications [2].

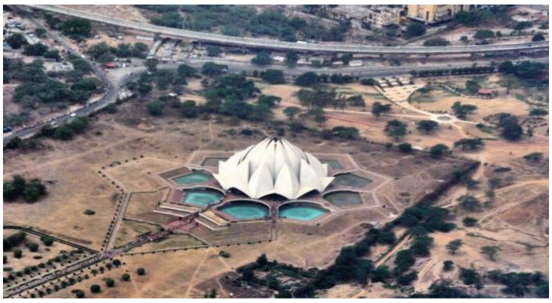

Fig. 1. Aerial View of Lotus Temple through PHANTOM 2 Pro UAV

1.2 Object Detection

Object detection is a challenging problem because objects can vary widely in appearance, size, orientation, and location within an image. In order to solve this problem, object detection algorithms typically use a combination of feature extraction, object recognition, and machine learning techniques to detect and classify objects [3]. Object detection is a computer vision task that involves identifying and localizing objects of interest within an image or a video. The goal of object detection is to determine the presence and location of one or more objects in an image or video and label them with specific category or class [4] [5]. In this research, the identification and localization of monuments is a primary goal through the techniques of object detection.

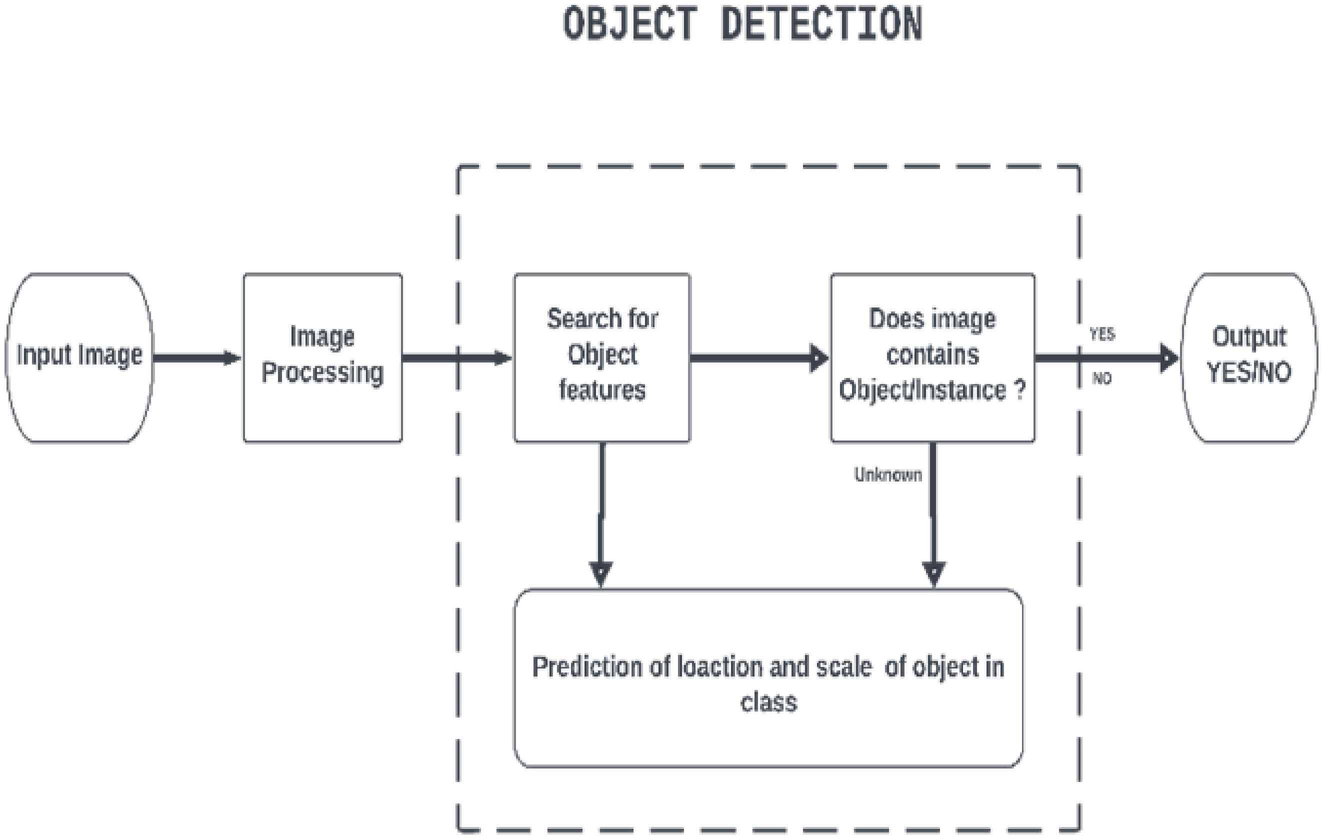
OBJECT DETECTION
Input Image
Image Processing
Search for Object features
Does image contains Object/Instance ?
YES
NO
Output YES/NO
Unknown
Prediction of loaction and scale of object in class

Fig. 2. Object Detection model

1.3 Heritage Sites and Monuments

India is ranked as sixth in number according to UNESCO cultural heritage Monuments sites. India is a country with a rich and diverse cultural heritage, and it is home to many monuments that are of historical, cultural, artistic, architectural, technical or political significance [6]. Delhi is the center of UNESCO heritage sites and our research spot for detection of some monuments like Humayun’s Tomb, Jama Masjid and Qutub Minar. Heritage sites helps us examine our history and traditions and enables us develop an awareness about ourselves [7].

Monuments are structures or landmarks that have been built to commemorate a person, event, or culture. They can be of historical, cultural, artistic, architectural, technical or political significance. Monuments are often erected as a way to remember important people or events, or to celebrate a culture or heritage. They can be a physical reminder of the past and can serve to educate and inform future generations about the history and culture of a place [8].

1.4 Computer Vision and Neural Networks

Computer vision has a vast expanded area of recording raw image data into the extraction of image patterns or features and interpretation of information. This includes the ideas of digital image processing, pattern recognition, computer graphics, image segmentation and artificial intelligence [9] (ex: detection of monuments in aerial images and localization in the images). Development of this filed is done by adapting the image processing (classification, detection and segmentation) and deep learning techniques [10].

Neural networks or artificial neural networks (ANNs)is a replica architecture of mammalian brain that perform the algorithmic calculation of data through layers of neurons and generate outputs in appropriate formats [11]. ANNs are a type of deep learning algorithm that is designed and work the same as the structure and function of biological neurons. They consist of layers of synchronously connected nodes, which process and transmit information. The input layer receives the input data, and the output layer predict the final output. The one or more hidden layers in between process the data and extract features. Each node receives input from the previous layer, applies a mathematical function to it, and sends the output to the next layer. The thresholds and weights of the nodes are fine-tuned during the training to optimize the network's performance on a specific task [12]. This research focused on implementation of Convolutional Neural Networks based algorithms.

II. METHODOLOGY

Implementation starts from the collection of aerial imagery dataset and labels the data and separate ground truth. Annotation of data is an important part of the process of implementation of the algorithm and training them for prediction on unknown data. From the literature survey, Finalize the two latest deep learning algorithms for this research.

1.5 Data Preparation

Data Collection is an important process for the preparation of dataset and a tough task to collect from various sources. Online collection of archaeological surveys through aerial drones, online photo galleries, and social media are the sources to collect aerial imagery data. Around 2000 aerial images were collected with resolutions from 240x240 to 2410x2410on different height scales.

1.6 Data Annotation

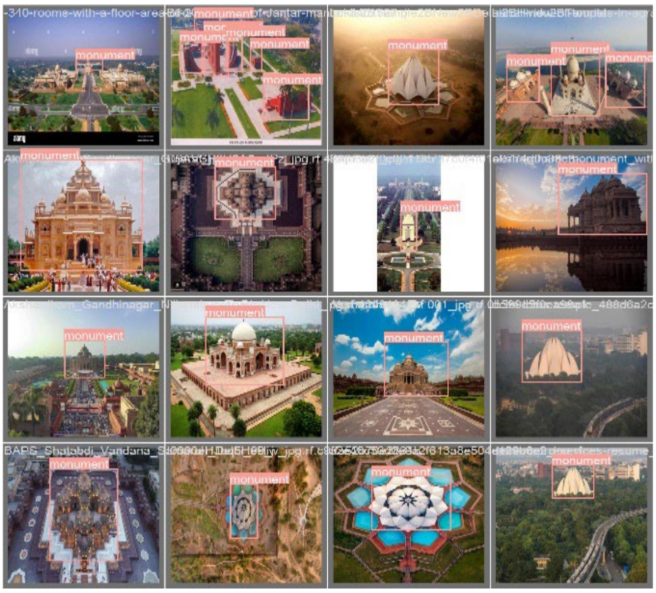

Fig. 3. Data Annotation for YOLOv5 and VGG-16

Annotation is the process of labeling the data to show the outcome of your deep learning algorithm. Labeling helps to train an algorithm on data and compare the accuracy after the model is trained. Label Img is one of the finest data annotation tools used to annotate aerial imagery data for the detection of monuments. It is easy to use and support different format (Yolo format) [13]. Labelling the data is possible through the bounding boxes annotation for detection and separates the labels and images file into training, validation, and test dataset.

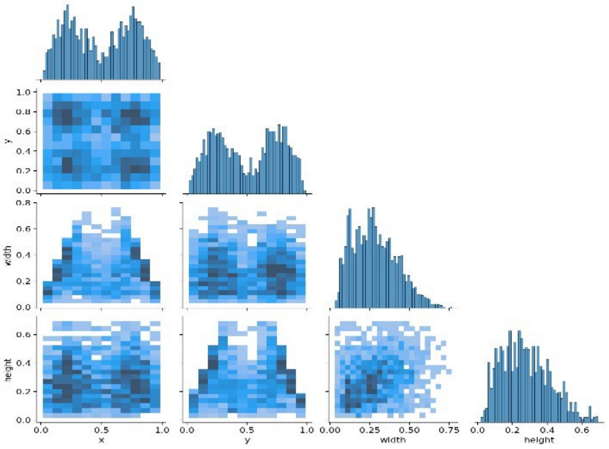

Fig. 4. Labels Correlogram

1.7 Data Augmentation

Data Augmentation gives diversity through various methodssuch as resizing, re-scaling, rotating, brightness scale, etc. It helps to create more data to train the model to overcome overfitting and improve the accuracy of the model. YOLOv5 introduced the mosaic augmentation technique to improve training efficiency. In Mosaic augmentation, four same or different images are randomly selected, resized, cropped, and combined as a single image. After this, four bounding boxes are placed in one mosaic augmented image which improves the model to detect small objects in the image [14].

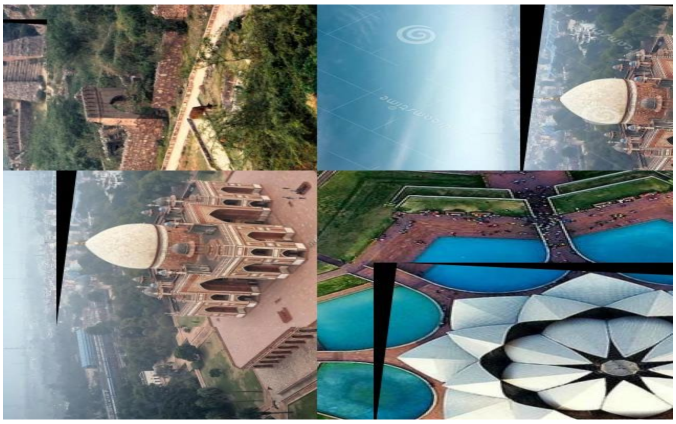

Fig. 5. Mosaic Data Augmentation

1.8 VGG-16

The visual Geometry Group of the University of Oxford developed the VGG-16, which is a convolutional neural network (CNN). It is known for its simple architecture and its ability to achieve high accuracy on image classification tasks. The model is composed of a mound of convolutional layers, where the number of filters increases as you move deeper into the network. The first convolutional layer (Conv-1) has 64 filters, the second (Conv-2) has 128 filters, the third (Conv-3) has256 filters, and the fourth and fifth (Conv-4 and Conv-5) have512 filters. After the stack of convolutional layers, there are three fully-connected (FC) layers. The first two FC layers have4096 channels each, and the third FC layer performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer of the VGG-16 model is the SoftMax layer, which is used for probability-based classification [15].

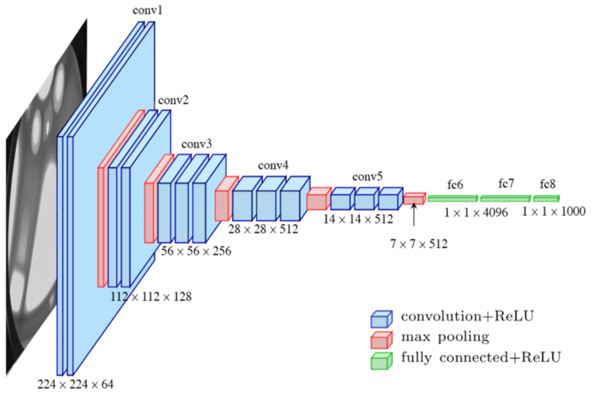

Fig. 6. VGG-16 Network Architecture

1.9 YOLOv5 PyTorch

YOLOv5 is a state-of-the-art object detection algorithm in the family of You Only Look Once (YOLO). YOLO v5 generally, uses a combination of different architectures to improve its performance. The backbone architecture used in YOLOv5 is CSPDarknet53, which is a variant of the Darknet architecture. It uses cross-stage partial connections (CSP) to improve the representation power of the model. The SPP layer is added to the CSPDarknet53 to upgrade the model's potential to handle instances of different scales[16]. The Neck architecture used in YOLOv5 is PANet (Path Aggregation Network) which helps to improve the feature pyramid and multi-scale context aggregation. The YOLO detection head is a convolutional neural network (CNN) architecture that is used to predict the bounding boxes andclass probabilities of objects in an image.

To further optimize the whole architecture, YOLOv5 provides a set of techniques called "bag of freebies". Bag of freebies are small modifications to the architecture that improve the performance of the model without adding additional computational cost. "Specials" are techniques that are added to the architecture to improve the performance of the model, but they may increase the computational cost. Therefore, it is a good choice for our research to implement a one-stage detector with good performance and convenience for implementation [17].

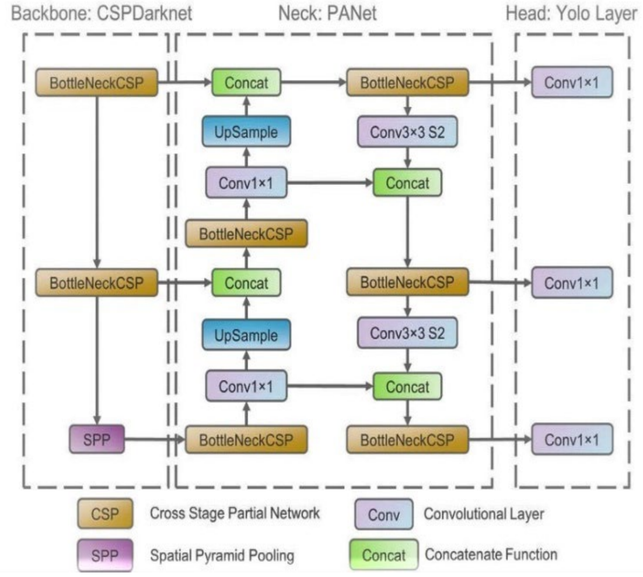

Fig. 7. YOLOv5 Network Architecture

III. RESULTS

The result was plotted of both models and the confidence score of 0.25 and learning rate of 0.001 was selected and then trained for 200 iterations, and accuracy of VGG-16 was raised to 63% and YOLOv5 was raised to 90%. Then saved models can be used for prediction, which is the process of applying the model to new data and using it to make predictions. The research project code repository on GitHub: https://github.com/ thegautamkumarjaiswal/ATEET

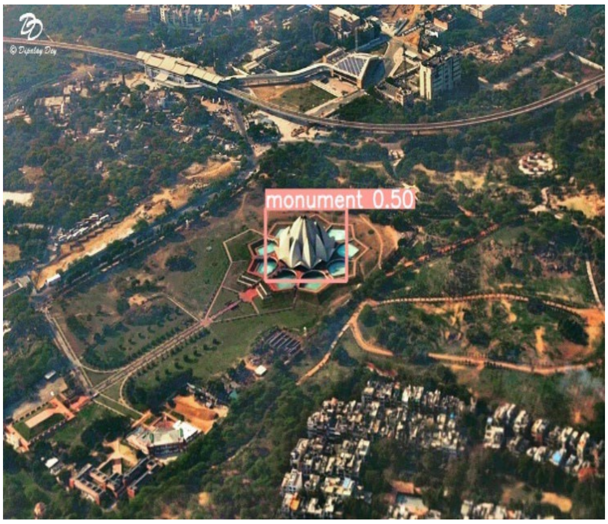

Fig. 8. Sample output 1 of VGG-16 model

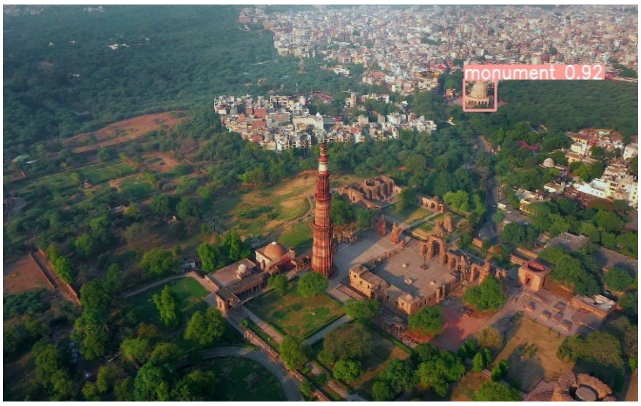

Fig. 9. Sample output 2 of VGG-16 model

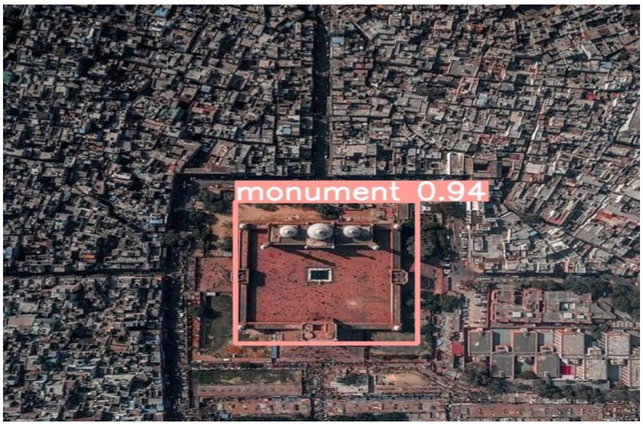

Fig. 10. Sample output 1 of YOLOv5

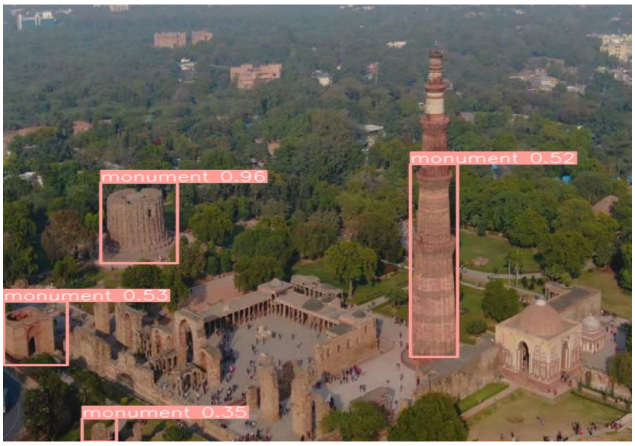

Fig. 11. Sample output 2 of YOLOv5

IV. EVALUATION METRICS AND ANALYSIS

For the VGG-16 model, data were split into training and validation with a ratio of 8:2 and trained for 200 epochs with a learning rate of 0.001, and batch size of 32 and this raised the accuracy to 63% as shown in Fig. 12. But the constant minor loss in training and noise loss at the time of validation as shown in Fig. 13 that shows the model is going to ward under fitting. The main reason behind the under fitting problem was having small size data. As compare to VGG-16, YOLOv5 performed better with this problem with expected accuracy.

YOLOv5 is the best model for object detection with less amount of data. With 200 epochs, the YOLOv5 model raised accuracy to 90% with other noticeable parameters such as mAP, Precision, Recall, and Loss as shown in Fig. 14.

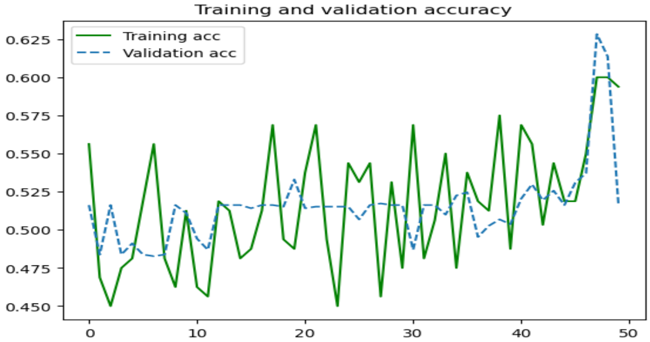

Fig. 12. Training and Validation accuracy of VGG-16 model

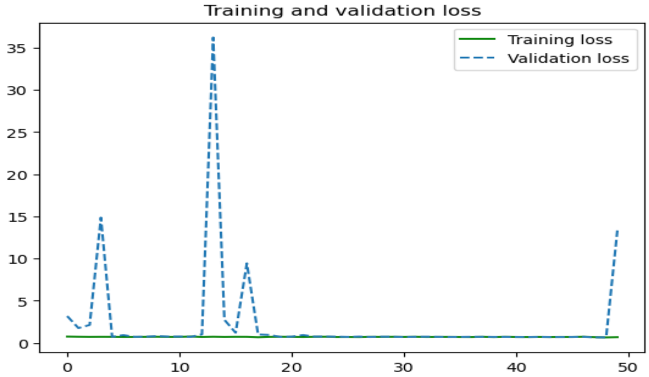

Fig. 13. Training and Validation loss of VGG-16 model

The Average Precision Mean (mAP), precision, and recall used to describe the experimental outcomes. To begin, we determine the precision rate and recall rate for each category of an object using the following formula.

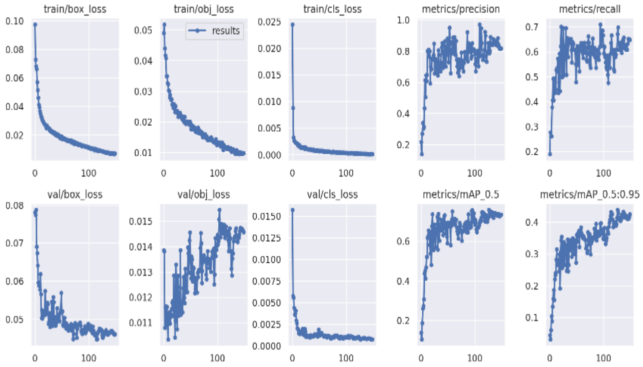

Fig. 14. Graph of training and validation losses, precision, recall and mAP
(Mean Average Precision with IOU 0.5 and 0.5:095)

Where TP (True Positive) and FP (False Positive); FN (False Negative); mAP is the average of all categories of AP, where AP (Average Precision) of a particular category.

$$\text{Precision}(\text{P}) = \frac{\text{TP}}{\text{TP+FP}}, \text{Recall (R)} = \frac{\text{TP}}{\text{TP+FN}}, \quad (1)$$

$$\text{AP} = \sum_{\text{n}}(\text{Recall}_{\text{n}} - \text{Rrecall}_{\text{n}-1})\text{Precision}_{\text{n}} \quad (2)$$

We evaluate the mAP averaged for IoU∈ [0.5:0.05: 0.95]. For every bounding box, we measure the overlap between the predicted bounding box and the ground truth bounding box. This is measured by IoU (intersection over union) [18].

V. CONCLUSION

This work implements VGG-16 and YOLOv5 to Monuments Detection, showcasing the architecture of both VGG-16 and YOLOv5 network architecture, process hyper-parameters, and dataset pre-processing. In the end, a comparison of both models through parameters such as training and validation losses, precision, recall, and mAP with IoU threshold. YOLOv5 found the best model compare to VGG-16 for Monuments Detection which can detect some mentioned monuments rapidly from aerial imagery. Using flask, we developed a web application named as ATEET - Monument Detector that allow the user to easily upload an image and display the result of the detection in an intuitive manner.

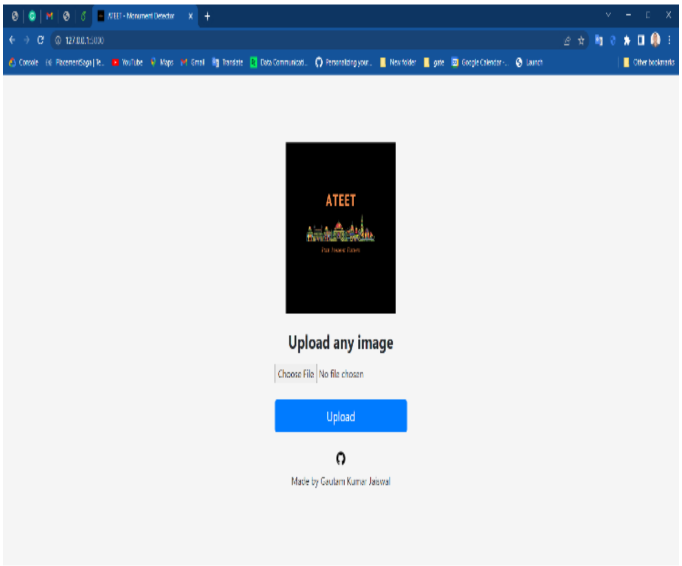
ATEET - Monument Detector
127.0.0.1:5000
ATEET
Upload any image
Choose File No file chosen
Upload
Made by Gautam Kumar Jaiswal

Fig. 15.ATEET - Monument Detector web application

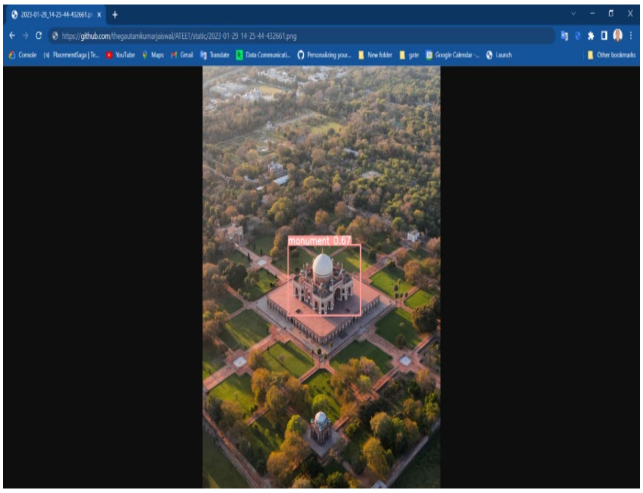

Fig. 16. Detection of monument through web application

VI. REFERENCES

[1]. Kartik Umesh Sharma and Nileshsingh V. Thakur, "A review and an approach for object detection in images," International Journal of Computational Vision and Robotics, vol. 7, pp 196-237, January 2017.

[2]. I.K. Nikolos, K. P. Valavanis, N. C. Tsourveloudis, and A. N. Kostaras, "Evolutionary algorithm based offline/online path planner for UAV navigation," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 33, no. 6, pp. 898–912, Dec. 2003.

[3]. S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 11, pp. 1475–1490, Nov. 2004.

[4]. N. Ghoggali, F. Melgani, and Y. Bazi, "A multiobjective genetic SVM approach for classification problems with limited training samples," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 6, pp. 1707–17, 18, Jun, 2009.

[5]. Thomas Moranduzzo and Farid Melgani, "Automatic Car Counting Method for Unmanned Aerial Vehicle Images" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 52, NO. 3, MARCH 2014.

[6]. Girshick, R. B. Donahue, J. Darrell, T. and Malik, J. Rich "Feature hierarchies for accurate object detection and semantic segmentation". CoRR, abs/1311.2524v5, 2014. Published in Proc. CVPR, 2014.

[7]. Mohd Anul Haq, Gazi Rahaman, Prashant Baral, Abhijit Ghosh, "Deep Learning Based Supervised Image Classification Using UAV Images for Forest Areas Classification" Journal of the Indian Society of Remote Sensing (March 2021) 49(3):601–606.

[8]. Prakash PilinjaSubrahmanya, Bharath Haridas Aithal, Satarupa Mitra, "Automatic Extraction of Buildings from UAV-Based Imagery Using Artificial Neural

Networks" Journal of the Indian Society of Remote Sensing 6 October 2020 / Accepted: 22 October 2020.

[9]. O.E. Apolo-Apoloa, Mart´ınez-Guantera, Egeaa, Rajab, P´erez-Ruiza, "Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV" European Journal of Agronomy 115 (2020) 126030.

[10]. Csillik, Cherbini, J. Johnson, R. Lyons, A. Kelly, "Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks". Drones 2, 39.

[11]. Altaweel, A. Marsh, S. M¨uhl, O. Nieuwenhuyse, K. Radner, K. Rasheed, and S. A. Saber, "New Investigations in the Environment, History, and Archaeology of the Iraqi Hilly Flanks: Shahrizor Survey Project 2009–2011." Iraq 74: 1–35.

[12]. Thomas Moranduzzo and Farid Melgani, "Detecting Cars in UAV Images with a Catalog-Based Approach" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 52, NO. 10, OCTOBER 2014.

[13]. Rachna Pathak, Anil Saini, Arnav Wadhwa, Himanshu Sharma, Dhiraj Sangwan, "An object detection approach for detecting damages in heritage sites using 3-D point clouds and 2-D visual data" Journal of Cultural Heritage 48 (2021) 74–82.

[14]. RaheeWalambe, Aboli Marathe and Ketan Kotecha, "Multiscale Object Detection from Drone Imagery Using Ensemble Transfer Learning" Detection from Drone Imagery Using Ensemble Transfer Learning. Drones 2021, 5, 66. https://doi.org/10.3390/drones5030066 Academic Editor: Anastasios Dimou Received: 15 June 2021 Accepted: 21 July 2021 Published: 23 July 2021.

[15]. Mariusz Wisniewski, Zeeshan A. Rana and Ivan Petrunin, "Drone Model Classification Using Convolutional Neural Network Trained on Synthetic Data" Classification Using Convolutional Neural Network Trained on Synthetic Data. J. Imaging 2022, 8, 218.

[16]. Hyun-Ki Jung and Gi-Sang Choi, "Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions" Academic Editors: Mauro Lo Brutto, Junchi Yan and Minghao Guo Received: 19 June 2022 Accepted: 17 July 2022 Published: 19 July 2022.

[17]. Oivind Due Trier, JarleHamarReksten, Kristian Løseth, "Automated mapping of cultural heritage in Norway from airborne lidar data using faster R-CNN" International Journal of Applied Earth Observations and Geoinformation 95 (2021) 102241.

[18]. L. N. Smith, "A disciplined approach to neural network hyperparameters: Part 1 – learning rate, batch size, momentum, and weight decay," 3 2018. [Online]. Available: http://arxiv.org/abs/1803.09820.