Neural Networks

Object Detection in Satellite Images Using Custom CNN

Team Members:

- Farag Esam Farag Abouelezz. 1000167006 Section5
- Kareem Mohamed Abdelkader Mohamed. 1000166627 Section5
- Fekry Ibrahim Megahed Megahed. 1000166586 Section5

Abstract:

Satellite image analysis is being increasingly used for many applications like surveillance, military, geo-spatial surveys and environmental impacts and change monitoring. Automatic detection and classification of objects is an important functionality of satellite image analysis. Due to the nature and size of objects and the varied visual features, it becomes challenging to detect and classify objects in aerial images. Manual detection of objects in these images is very time? consuming due to the nature and that data captured in these images. It is desirable to automate the detection of various features or objects from these satellite images. The conventional methods for object classification involve two stages: (i) identify the regions with object presence in the image and (ii) classify the objects in the regions. Additionally, detection of objects becomes challenging in presence of complexities in background, size, noise, and distance parameters. This work proposes a customized convolutional neural network to detect and classify three different objects such as trees, building and cars in the images. It also aims to understand and outline briefly the performance characteristics of the considered custom CNN.

CNN (Convolution Neural Networks).

Keywords— Satellite imagery, object detection, classification, custom, convolutional neural networks, image processing

Problem Definition (or Motivation):

Object detection is a subsection of computer vision and refers to the process of determining a class or category to which an identified object belongs to and estimating the location of the object by generating a bounding box around it. Deep Convolutional Neural Networks (CNNs) have been used extensively for object detection and have consistently achieved remarkable results. Thanks to the availability of large datasets, powerful computing resources, and constant innovation around network architectures, increasingly faster and more accurate models surface every day.

So, what exactly is object detection used for? It is applicable to many domains across industries including surveillance (tracking entities like people, vehicles or identifying unattended baggage), medical diagnosing (detecting lung nodes or localizing lesions), or transportation (detecting objects on the road for autonomous driving). While these are more specialized use cases, the value of object detection more generally lies in the fact that information can be obtained in a much faster and much more accurate way and at a significantly reduced cost than with conventional data collection methods.

Satellite imaging is gaining importance in many applications like remote surveillance, environmental monitoring, aerial survey etc. All these applications involve searching objects, event of interest, facilities etc., from the satellite images. In most applications, manual detection and classification of objects becomes very difficult especially with large volumes of data and the number of satellite images to process collectively. Though detection and classification of objects in images is a long-studied topic in image processing domain, detection in satellite (aerial) images is more challenging as the objects are small and their visual features are extremely hard to track and capture making it all the more difficult. Towards this end various automated techniques for detection and classification have been proposed and are in the works. From classical machine learning (ML) to current deep learning, many solutions have been proposed for object detection and classification in satellite images. Out of these methods machine learning methods for detection and classification are the most researched over the last few decades. These methods involve extraction of various features from the images and classifying them using ML classifiers. Automated Object detection is still a challenge due to variation in the size of the object, orientation, and background of the target object. Conventional machine learning classifiers involving manual selection of features like HOG, Gabor, Hough transform, wavelet coefficients etc., are not able to address the challenges in automated object detection. Hence, there is a need for an efficient approach, and Deep Learning has shown promising results in achieving the objective of detection and classification using CNN. Recently the Deep learning classification methods which have been proposed for automated object detection with high accuracy are able to learn features automatically from the images instead of manual selection of features. Many deep learning models based on convolutional neural network (CNN) are proposed for detection and classification of objects in satellite images. These models involve two steps. In the first step, the regions of presence of object in the image are detected. In the second step, the objects are classified using convolutional neural network.

In this work, a customized convolutional neural network is proposed to detect and classify objects in satellite images. The model is trained to classify three objects of trees, building and cars in the satellite images. The detection and classification performance are compared with actual execution of YOLO V3 algorithm for the same dataset and some standard benchmark data for other algorithms without execution though. YOLOV3 algorithm combines the detection and classification of objects in a single stage instead of two passes as done in conventional CNN models.

RELATED WORKS:

The existing deep learning models for detection and classification of objects are surveyed in this section. In

[1] the authors have proposed a variation using a CNN to detect objects. For satellite images specifically they put forth the concept of a rotation invariant region based convolutional neural network. In that before classification of the objects begins the step of normalization of feature representation is taking place to achieve and focus on the concept used of a region (rotation invariant). After the same then classification is carried out which is based now on the fact that for each image there is a higher complexity on computation for patching results through rotation invariant regions for each image.

Authors in [2] are outlining a method for detection of bridges and large crossovers on bodies of water which are shown in satellite images. Water bodies primarily here are rivers which need to be detected or recognized firstly in the image by a technique called as recursive scanning which uses geometric constraints for identifying such details like it is a river type of water body. Thereafter using these identified rivers in the first step, and on the basis of application of the knowledge relate to spatial dimensions concerning different bridges a scan is performed over the extent of the identified rivers to further detect or identify the pixels which could be belonging to a bridge. Once the pixels are identified then an analysis is performed as to the relationship or connectivity of the identified pixels and based upon that analysis it is determined whether a bridge segment is identified in the image for the pixels which got identified first. Although one problem exists for this kind of an approach and that is that we need to have a prior knowledge and understanding of the spatial dimensions of the objects or structures we are trying to identify, which in this case is bridges.

Authors in [3] have detailed a two staged approach for detecting networks of road that are seen or exist in aerial images like ones taken from satellite or UAVs. It leads to automatic detection through first the detection stage and then cutting down or pruning stage being applied. A Bayesian model is used first for classification of shapes of regions which are homogeneous or similar characteristic regions or shapes by detecting these in this stage. The second stage comprises of ascertaining the likelihood of any particular part or segment being a road through the use of a technique called conditional probability. This technique like many others used in detection is very high in its computational complexity.

Authors In [4] the authors propose an image analysis technique which is object based and is used in finding the land cover and its primary usage for classification of the topology of different stretches of areas in the satellite images. Before an SVM classifier can be trained we need to extract texture features and for that all the objects from the image are first segmented. Once this is done, then these segmented objects are used for carrying out the action of extraction. Once the segmented objects' texture features are extracted it is used to train an SVM classifier for the purpose of classifying the land based on its cover and usage. Relevant to this scheme there is a limitation on accuracy due to feature extraction, which is normally constrained and not exhaustive or very complete or accurate in itself.

EVALUATION OF EXISTING WORK:

There are multiple implementations in the field of AI and Deep Learning with which object detection can be performed. But each of those methods have their own advantages and disadvantages. In this section, the most popularly used implementations such as Faster R-CNN, ResNet and YOLO V3 has been selected. Furthermore, they will be evaluated based on performance metrics namely accuracy, precision, recall and F1 score. In Table 1, the above-mentioned comparisons have been represented.

Table 1 : PERFORMANCE EVALUATION METRICS OF EXISTING SYSTEMS

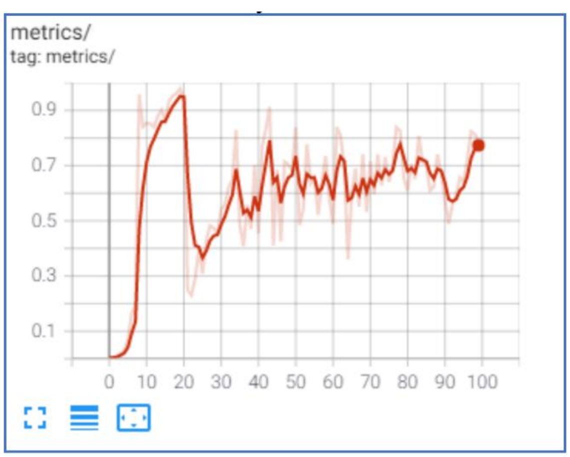
Implementation	Accuracy	Precision	Recall	F1 score
YOLO V3	90.40	0.90	0.89	0.89
ResNet	87.23	0.87	0.86	0.86
Faster R-CNN	83.64	0.83	0.82	0.82

PROPOSED METHODOLOGY:

- Image Acquisition and data processing.
- Preparation of data and preprocessing.
- CNN construction and training.
- CNN validation and analysis of results.
- Drawing of bounding box.

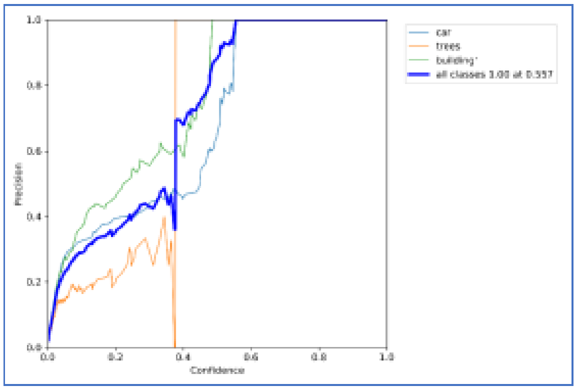
System during run-time (with 50 epochs)

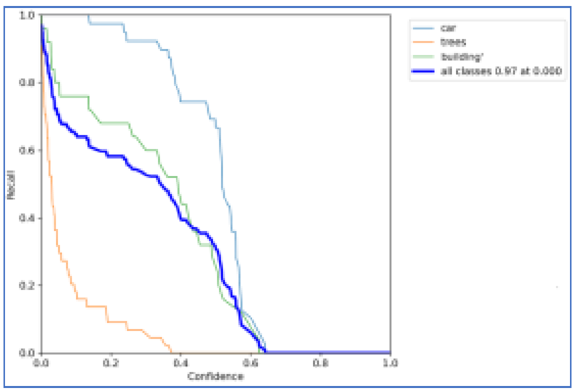
System during run-time (with 100 epochs)

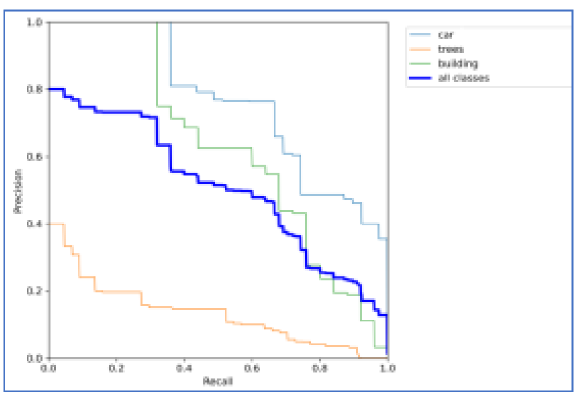


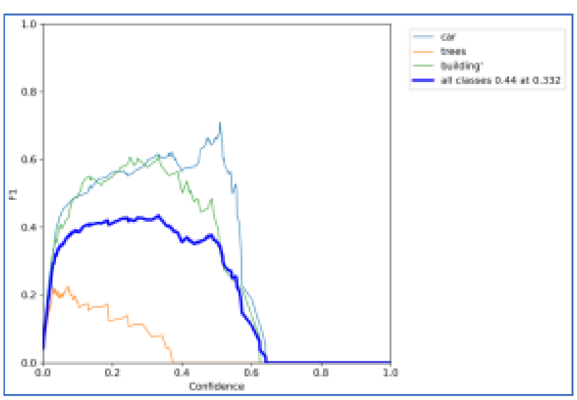
the final evaluation metrics have been represented with the accuracy of 94.65%.

Results:







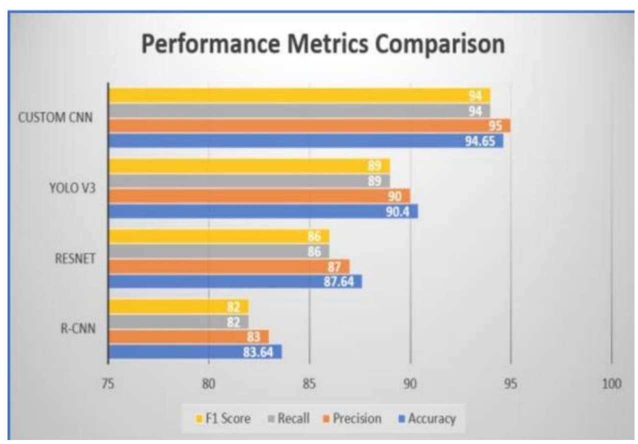


A better balance between precision and recall is achieved for car object followed by building and trees in the proposed solution:

• Precision plot over different confidence - Car objects are classified with higher confidence in the proposed custom CNN, followed by buildings and trees.

• Ration between precision and recall - The recall is higher for car followed by building and tress objects in the proposed custom CNN solution

. • F1 measure plot for different class of objects - From the results, car and building objects are classified far better compared to trees in the proposed solution. After successful detection, a comparison graph was plotted to compare the proposed custom convolutional neural network-based implementation to the other implementation mentioned in the comparison, in Figure 10 below. From comparison, we can clearly observe the edge the custom CNN has over the other implementations for detection of objects from satellite imagery. Furthermore, YOLO v3 and 4 were tested extensively for the same use-case scenario, but it yielded just around 90% net prediction accuracy. The proposed custom CNN solution has 4.25% more accuracy compared to YOLO v3.



References:

[1] Detection and Classification of Objects in Satellite Images using Custom CNN – IJERT

[2] Object Detection With Deep Learning on Aerial Imagery (dataiku.com)

[3] Object detection on Satellite Imagery using RetinaNet (Part 1) — Training | by Ijeoma | Medium

[4] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. CSPNet: A new backbone that can enhance learning capability of cnn. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop), 2020.

[5] Bochkovskiy, Alexey & Wang, Chien-Yao & Liao, Hong-yuan. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection., 2020.

[6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.