RESEARCH ARTICLE OPEN ACCESS

Leveraging Transfer Learning for Precise Geospatial Object Detection

Emmanuel Chinyere Echeonwu*, Moses Okechukwu Onyesolu**, Bolou Dickson Bolou***
*(Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
Email: ec.echeonwu@stu.unizik.edu.ng)
** (Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
Email: mo.onyesolu@unizik.edu.ng)
***(Department of Computer Science, Nigeria Maritime University, Okerenkoko, Delta State, Nigeria
Email: bolou.boluo@nmu.edu.ng)

--************************----------------------------------

Abstract:

Object detection in geospatial imagery plays a critical role in various applications, including land-use analysis, environmental monitoring, and disaster response. Traditional computer vision and image processing techniques employed for object detection can be labor-intensive and time-consuming, hindering the efficiency of processing geospatial data, particularly for real-time applications. Deep learning, however, offers a powerful alternative with its ability to automate feature extraction and achieve superior classification performance. This study investigates the application of Faster Region-based Convolutional Neural Network (Faster R-CNN), a state-of-the-art deep learning model, for object detection in UAV imagery. The research methodology encompasses the acquisition of UAV data, followed by the careful annotation and preparation of a training and validation dataset encompassing diverse object classes. Annotations adhere to the PASCAL VOC standard, ensuring data quality and facilitating model training. The Faster R-CNN model is then employed for object detection within the prepared dataset. To assess the effectiveness of the trained model, a comprehensive validation process is undertaken, incorporating both qualitative and quantitative evaluation methods. The experimental results demonstrate a promising mean average precision (mAP) of 0.87 across a representative sample of test images, signifying the model's ability to accurately classify and localize objects within UAV imagery. These findings highlight the potential of deep learning techniques in empowering geospatial analysts with robust and efficient tools for visual recognition tasks. By automating feature extraction and achieving superior classification accuracy, deep learning paves the way for faster and more comprehensive analysis of geospatial data, ultimately contributing to more informed decision-making processes within the field.

Keywords **—Faster Region-based Convolutional Neural Network, Remote Sensing, Geospatial Intelligence, Image processing, Object detection, Unmanned Aerial Vehicle, Deep learning. Earth Observation.**

--************************----------------------------------

I. INTRODUCTION

Machine learning (ML), one of the most important subsets of artificial intelligence (AI) has transformed how data is processed; and are used to answer challenging questions. This is also true for the emerging geospatial industry where the use of digital technology to meet the fourth industrial revolution (4IR) has been the hallmark. Thus, AI and ML applications are of great importance in many spatial-enabled systems, which include, but not limited to, smart systems, transportation, precision agriculture, urban planning, disaster management, environmental monitoring and management, etc.

Driven by the usefulness of AI and ML, geospatial professionals have made tremendous progress over the years to improve image processing and interpretation through varied object detection methods [8]. This is because geospatial image processing and interpretation can be a difficult and daunting task to accomplish within the geospatial and remote sensing community. These challenges have been attributed to the enormous amount of data collected using ubiquitous geo-sensors which increases the spatio-temporal and spectral resolution of the data. Besides, data collected from these sensors exhibit heterogeneity and uncertainty [9]. However, with the advancement in ML, it has been found to outperform the human-led interventions in a timely, accurate and consistent manner. No doubt, deploying machine learning algorithms to these image processing tasks offer enormous advantages and improve the performance of the present geospatial image processing techniques (e.g. feature extraction, image enhancement, change detection, etc) [2].

Deep learning is a branch of machine learning which is propelled by the development of Artificial Neural Networks (ANNs). The neural networks as a class of the learning algorithms were inspired by the structure and functionality of the human brain, with its basic structure known as the neuron. For the neural network, a perception forms its basic structure. A single perceptron algorithm takes in some inputs which have been weighted and computes an output using activation function. Typically, a neural network algorithm is made up of one or more interconnected layers of perceptrons which are usually configured to accept some data inputs. A single layer perceptron describes one layer network and a multi-layer perceptron describes more than one interconnected layers. Neural networks are trained so that the weight parameters are optimized to find the best possible sets.

Indicated earlier, deep learning networks are part of the ANNs family with the depth in the layers depicting the deep aspect of the network. The deep learning networks can be employed as supervised, unsupervised and hybrid networks respectively. Unlike the unsupervised and hybrid form of the network, the supervised networks fully require that the task provides specific information about the datasets before the learning process [11].

The discovery of the convolutional neural networks (CNNs) gave the push needed to advance the research in object detection and continues to have great influence in the field of computer vision. The successful implementation of imagenet image recognition task enhanced the explosion in research and adaptation of different deep neural architectures for many computer vision tasks.

The CNN is an important part of a neural network because of the most vivid advantage of automatic feature extraction, hence the now adopted paradigm called end-to-end learning. To do this, the network uses kernels, also known as filters, to detect and extract features in the input image. Its basic operation involves the dot multiplication of the kernel and the overlap region, mostly referred to as the receptive field, in the image [11].

With the emergence of various spatial data recording systems, such as artificial satellite systems and drones etc, deep learning algorithm has been implemented through transfer learning to bring insights to reality where large processing capability of the graphical processing unit (GPU) is an enabling factor. Nonetheless, this cannot be said without the inevitable challenges of dealing with geospatial data which do not go through the most conventional process of data capture [2]. Often, data is recorded at a certain angle which can make information recorded not immediately recognizable. Another obvious challenge is the non-orderliness and non-equal representation of geospatial objects at any point in time. This creates room for a lot of ambiguity and under-representation.

II. BACKGROUND

To successfully build a deep learning model, one major obstacle is the large-scale data needed for learning. This is also in addition to the fact that the training data and test data has to be independently and identically distributed (iid). However, it is often inevitable to encounter insufficient data and data collection has become almost complex and expensive [14].

Transfer learning relies less on the assumption that the training data must be independently and identically distributed (iid) with the test data and the model does not need to be trained from scratch. This reduces greatly the need for enormous training data and time [14].

Four categories of deep transfer learning namely instance-based, mapping-based, network-based and adversarial-based approaches could be implemented. Li et al. defined the instance-based approach as one which uses specific weight strategy and utilizing some instances from the source domain. [7]. It builds on the assumption that there maybe differences arising from the source domains, some instances can be shared with a target domain with carefully selected weights. Long at al. defined a mapping-based approach as a way of mapping instances from both the source and target domains into a new data space [10]. This approach assumes that data can originate from a different domain but there could be some similarities when mapped to new data space. While network based approach uses a partial network that has been pre-trained on the source domain, including its network structure and parameters as part of the neural network in the target domain. The adversarial based approach introduces the adversarial methods originally motivated by the use of generative adversarial networks and finds transferable representation between the source and target domains [1][12]. In this study, the deep learning transfer approach based on instance strategy was implemented.

Object detection is an important aspect of computer vision applications. While the fundamental task in computer vision started with image recognition, object detection seeks to further ascertain the concepts of locations of objects in the image [3].

As a fundamental task in computer vision, object detection aids image understanding connotatively and in many ways, can be related to many other application such as image classification, face recognition, object tracking and autonomous driving [15]. Object detection methods falls into two categories. The first category proposes the region-based methods and then classifies those proposals into different object categories. The second category applies a single stage method and treats the object detection as purely a classification or regression problem [15].

TABLE I
UNIT AND INTEGRATION

S/No	Region-Based Methods	Single Stage Methods
1	R-FCN	MultiBox
2	SPP-Net	AttentionNet
3	FPN	G-CNN
4	R-CNN	DSOD
5	Fast R-CNN	SSD
6	Faster R-CNN	DSSD
7	Mask R-CNN	YOLO

This study applied the faster R-CNN network for the object detection. The implemented approach is currently considered state-of-the-art in the region-based category. The faster R-CNN model modifies the fast R-CNN network proposed by Girshick [4]. The technique implements the Region Proposal Network(RPN) to replace the selective search method previously used in Ren et al [13]. It is noteworthy that the region based framework performs object detection tasks in three steps: feature extraction, region proposal, and classification and localization. Zhong-Qiu et al., defined the RPN as a fully convolutional network and has the ability to generate a set of object proposals from an image. This is achieved by sliding over the feature maps from the base network. This result is fed into the two equally related fully connected layers, class layer and the regression box layer, for the prediction of the object category and the bounding boxes respectively[15]. To fit the model, the loss function is given in Equation (1) as:

$$L(p_i,\mathrm{t}_i) = \frac{1}{N}\sum L_c\left(p_i,\mathrm{p}_j\right) + ƛ\frac{1}{N_r}\sum p_j\, L_r\left(t_i,\mathrm{t}_j\right) \qquad (1)$$

where p_i is the predicted probability of an object. The ground truth p_j is 1 if the object is found and 0 if it is not found. t_i is the parameterized coordinates of the predicted bounding box while t_j is the ground truth box overlapping the predicted object. Therefore L_c is a binary log loss while L_r is a smoothed regression loss used to fit the bounding box.

Different methods have been employed to evaluate the performance of object detection models. In recent times, with the emergence and adoption of the CNNs enhanced with the GPU-accelerated deep-learning frameworks, object- detection algorithms are currently being developed and measured. Detection algorithms such as R-CNN, Fast R-CNN, Faster R-CNN, R-FCN, SSD and Yolo have arguably accelerated the performance standards on these tasks.

With the object detection model trained, the performance should be verified. For classification task, models are only evaluated by computing the probability of the object class seen in the image. In this case, it is a simple task for the model to easily identify predictions that are correct from ones that are not. However, the object detection task extends further by localizing the object with a bounding box which is measured with a corresponding confidence score to show how certain the predictions are made.

One important performance metric used for evaluating the performance of object detection models has been the mean Average Precision (mAP) sometimes referred to as the Average Precision (AP). The mAP is defined as the metric used to measure the accuracy of prediction of a detection model with comparison to the ground truth annotated dataset. However, literature has revealed that different results suggest that choosing a better model architecture and weights undoubtedly go beyond considering only the mAP metric. Different features have been identified to contribute to a good performance such as bounding box tightness (IoU), high confidence false positives, individual spot performance and how the model performed at task more important [5].

$$\mathrm{mAP}=\frac{1}{N}\sum_{\mathrm{i}=1}^{N}\mathrm{AP}_i(2)$$

where AP_i is the average precision over all the classes.

$$\mathrm{AP}=\frac{1}{11}\sum_{\mathrm{r}=0,0.1,0\ldots0.9,1}^{n}P_i\,(r)(3)$$

where $P_i(r)$ is the interpolated precision used to describe area under the Precision-Recall curve.

The IoU is derived to compute the mAP [6]. The IoU measures the overlap between the predicted and the ground truth of the object. IoU of 0 means no overlap and IoU of 1 means a complete overlap. The IoU is given in Equation (4) as:

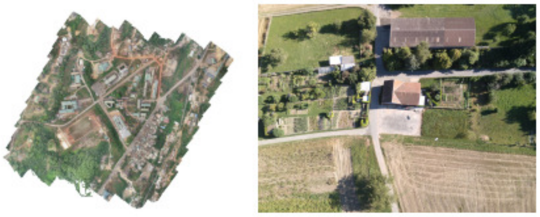
$$\frac{\mathrm{Area-of-overlap}}{\mathrm{Area-of-uunion}} = \frac{\mathrm{Prediction}}{\mathrm{GroundTruth}}(4)$$

Given that the IoU is an important accuracy metric, the best practice is usually to fix a minimum IoU requirement for various annotated objects. This ensures that for any annotations done, it is set to have IoU >= X where X = 0.95. Ironically, state-of-the-art detection systems do not perform at 0.95 IoU. These models have been reported to perform at less than one percent mAP (Fig. 1). Therefore atomic evaluation has been introduced and exploited to ascertain the ability of object detection models and more generally deep learning models for computer vision task [5]. This method becomes even more practical and realistic as it tends to look at the performance of the model on a case by case basis. In addition to the quantitative analysis, this allows for a broader insights of the strengths and weaknesses of the model prediction and reliability of the datasets such as prediction and ground-truth analysis, difficulty analysis, uniqueness analysis, redundancy analysis, annotation mistakes analysis etc, instead of relying on a single metric. In addition to the IoU, mean average precision (mAP) metrics, indicators such as precision, recall and F1 score were applied.

III. METHODOLOGY

Drone images covering different areas of interest were acquired. The images were tiled into a smaller area sizes. A total of 119 images were carefully selected as valid dataset for training, validation and testing phases. Out of the 119 images, 88 images were randomly selected as training data, 12 images were selected for validation dataset and 19 images were used to independently access the optimum trained model performance. The training and validation sets were drawn from the same distribution covering an area of interest while the test set was a collection of images covering both areas of interests. The datasets were preprocessed and given to the limited amount of training dataset, data augmentation was implemented. Figure 1 shows the areas of interest used for the study.

The datasets were fully annotated using the Pattern Analysis Statistical Computational Learning (PASCAL VOC) standard. This standard was developed by the European Union, and has been adopted as a dataset format in Extended Markup Language (XML) file format used for various Visual Object Challenge from 2005 till 2012. The number of annotations and objects distribution are shown in Figure 2. The following eight (8) classes were identified as object for detection: house, car, ongoing construction, built-up area, water storage, bus, truck, train track. The Faster R-CNN was used and the model initialized with pre-trained weights of the previously trained model. The Faster R-CNN is a state-of-the-art detection algorithm and it is made up of the resnet50 backbone network and the Region Proposal Network. Outputs from these networks in turns are fed into the classifier layer and the regression layer for the final predictions of class category and location bounding boxes. The model was trained with the dataset and optimized for better weight parameters. The results were analyzed to ascertain the performance of the model.



(a) Site 1. (b) Site 2.

Fig. 1 Areas of interest

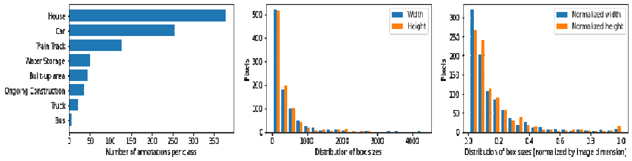


Fig. 2 Distribution of objects and annotations

IV. RESULTS AND DISCUSSION

In this study, both the qualitative and quantitative analyses of the Faster R-CNN model performance were carried out. Image uniqueness measure was applied on the test images to ascertain the performance of the model (Fig. 4). This image uniqueness score measures how similar the data are and can be used to determine duplication of data. It is usually measured on a scale of [0,1], with higher values indicating the best. From Figure 3, it can be observed that the uniqueness scores of the sampled test images were > 0.6.

Fig. 3 Sample test images

Qualitative analysis was performed to show the visual representation of model performance . Fig. 4 shows model's output and geo-objects detected. it is important to state that the model was able to detect object of centimeter resolution such as cars and rail tracks. The model also performed well on images with different camera configurations which were not originally used in the training sets. Figure 5 shows visual comparison of the test images of the ground truths and model predictions. Thus, Figure 5 highlights positional object detection with ground truth data. From the image superposition scheme that was done, it can clearly be seen in Figure 5 that the model's prediction are in agreement with the ground truth information.

In modeling, quantitative analysis is important because it provides the accuracy measure using appropriate metric for model performance. In this study, computing the model performance metrics was done using four(4) randomly selected images from the test dataset. Tables 2 to 4 show the true positives (TP), false positives (FP) and false negatives (FN) at different IoU thresholds. For this study, the IoU metric, precision, recall, Average Precision metric (AP) and mean Average Precision metric (mAP) for each test sample image set were utilized. IoU threshold values of 0.30, 0.50 and 0.70 were experimented to compute the true positives, false positives, and false negatives. However, threshold of 0.5 was adopted as the standard IoU for this work.

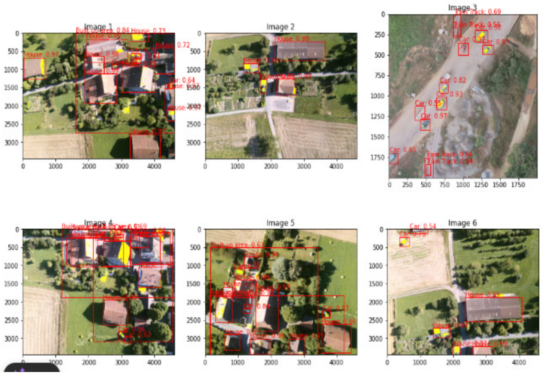
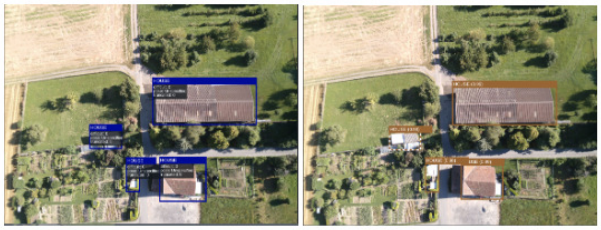


Fig. 4 Classification and localization

TABLE II
TP, FP, FN METRIC WITH IoU = 0.5

TRUE POSITIVES	FALSE POSITIVES	FALSE NEGATIVES
4	7	0
4	4	0
11	60	0
17	16	0



(a) Ground truth. (b) Model output.

Fig. 5 Ground truth vs model output

11-point average precision is an evaluation method used to evaluate how well an object detection system has categorized a set of detection. IoU threshold of 0.5 was selected for the computation. The precision values can be interpreted as the measure of the exactness of classification after prediction. The recall, on the other hand, is communicating the model's ability to detect positive instances by measuring the fraction of positive instances that are correctly classified. The F1 Score depicts the overall model performance which is categorized in the range of zero to one, with high values indicating high classification performance and vice versa. At the end, a mean average precision (mAP) of 0.87 was obtained by finding the average of the reported average precision results for the images in Table III.

TABLE III
MAP OF TEST IMAGES

	Mean Average Precision	Total average
Image 1	0.9	
Image 2	0.9	
Image 3	0.79	
Image 4	0.89	0.87

V. CONCLUSIONS

This study demonstrates the application of transfer learning techniques using deep learning to implement object classification and localization. A state-of-the-art Faster R-CNN model was developed and used to show that it is able to classify and localize geo-objects on aerial imagery. The work was done from bottom-up using raw aerial images captured at different time and locations with varying sensor settings and atmospheric conditions depicting different spatio-temporal characteristics. Images were fully custom-annotated with objects of interest. Thus, this work is able to show that in transfer learning, the dataset does not have to be independently and identically distributed. This study also affirms that computer vision tasks can be effectively performed with deep learning with impressive accuracy. This will encourage the use of deep learning for object detection tasks using geospatial data and will help improve spatial analysis efficiency.

REFERENCES

[1] Ajakan, H., Germain, P., Larochille, H., Laviolette, F., & Marchand, M., (2014). Domain-adversarial neuralnetworks. ArXiv preprint arXiv: 1412.4446. 10.

[2] Asokan, A., Anitha, J., Ciobanu, M., Gabor, A., Naaji, A., & Hemanth, D.J. (2020). Image processing techniques for analysis of satellite images for historical maps classification—an overview. Applied Sciences, 10(12), p.4207.

[3] Galvez, R., Bandala, A., Dadios E., Vicerra R., & Maningo, J. M. (2018). Object detection using convolutional neural networks. Proceeding of TENCON 2018. IEEE Region 10 conference.

[4] Girshick, R. (2015) . Fast R-CNN. ArXiv: 1504.08083.

[5] Hofesmann, E. (2020). IOU a better detection evaluation metric. https://towardsdatascience.com/iou-a-better-detection-evaluation-metric-45a511185be1.

[6] Hui, J. (2018) . mAP (mean Average Precision) for object detection. https:////medium.com/@jonathanhui/map – mean – average – precision – f or – object – detection –45c121a31173.

[7] Li, N., Hao, H., Gu, Q., Wang, D., & Hu, X. (2017). A transfer learning method for automatic identification of sandstone microscopic images. Computers and Geosciences. 103 (111-121).

[8] Li, K., Wan, G., Cheng, G., Meng, L., & Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159, pp.296- 307.

[9] Li, Z., Gui, Z., Hofer, B., Li, Y., Scheider, S., & Shekhar, S. (2020). Geospatial information processing technologies. In Manual of Digital Earth (pp. 191-227). Springer, Singapore.

[10] Long, M., Lao, Y., Wang, J., & Jordan, M. (2015). Learning transferable features with deep adaptation networks. International Conference in Machine learning.

[11] Mahony, N. D., Campbell, S., Carvalho, A., Harapanahalli, S., Hernadez G., Krpalkova, L., Riordan, D., & Walsh, J. (2019). Deep learning vs Traditional Computer Vision. ArXiv: 1910.13796.

[12] Oguab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representation using convolutional neural networks. Computer Vision and Pattern Recognition (CVPR). IEEE.

[13] Ren, S., He, R., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal network. ArXiv: 1506.01497.

[14] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Kiu, C. (2018). A survey on deep transfer learning. ArXiv: 1808.01974.

[15] Zhong-Qiu, Z., Peng, Z., Shou-tao, X., & Xindong, W. (2019). Object detection with deep learning: A review. IEEE Transaction on Neural Networks and learning systems.