TẠP CHÍ KHOA HỌC ĐẠI HỌC HẠ LONG
Scientific Journal of Ha Long University
http://uhl.edu.vn/

A REVIEW OF DEEP LEARNING-BASED ALGORITHMS FOR OBJECT DETECTION IN SATELLITE IMAGES

Nguyễn Trung Hiếu[1*]

[1]*Khoa Toán – Tin học và Ứng dụng Khoa học và Công nghệ trong Phòng chống tội phạm, Học viện Cảnh sát Nhân dân*

** Email: hieunt.dcn@gmail.com*

Ngày nhận bài: 03/10/2024 *Ngày nhận bài sửa sau phản biện: 09/12/2024* *Ngày chấp nhận đăng: 15/12/2024*

ABSTRACT

Object detection in satellite images is a particularly interesting area in computer vision. This paper synthesizes and analyzes the challenges and characteristics of satellite images, as well as existing methods, with a special emphasis on the role of deep learning. The authors point out that object detection in satellite images is different from that in conventional images due to the high resolution, noise, and diversity of objects. To address these challenges, this paper introduces anchor-based and non-anchor-based methods in detail, and highlights the advantages and disadvantages of each method. In particular, the emergence of Transformer architectures in computer vision has opened up a new promising direction for object detection in satellite images. In addition, this paper also discusses practical applications of object detection in satellite images, including environmental monitoring, resource management, and disaster response. Finally, the paper suggests potential future research directions, such as developing more efficient models, handling small objects, and leveraging diverse data sources.

Keywords: *computer vision, deep learning, object detection, satellite imagery.*

NGHIÊN CỨU CÁC THUẬT TOÁN HỌC SÂU TRONG PHÁT HIỆN ĐỐI TƯỢNG TRÊN ẢNH VỆ TINH

TÓM TẮT

Vấn đề phát hiện đối tượng trong ảnh vệ tinh đang là một lĩnh vực được quan tâm đặc biệt trong thị giác máy tính. Bài báo này tổng hợp và phân tích các thách thức, đặc điểm của ảnh vệ tinh, cũng như các phương pháp hiện có, đặc biệt nhấn mạnh vai trò của học sâu. Các tác giả đã chỉ ra rằng, phát hiện đối tượng trong ảnh vệ tinh khác biệt so với hình ảnh thông thường do độ phân giải cao, nhiễu và sự đa dạng của các đối tượng. Để giải quyết những thách thức này, bài báo đã giới thiệu chi tiết các phương pháp dựa trên anchor và không dựa trên anchor, đồng thời làm rõ ưu nhược điểm của từng phương pháp. Đặc biệt, sự nổi lên của kiến trúc Transformer trong lĩnh vực thị giác máy tính đã mở ra một hướng đi mới đầy hứa hẹn cho việc phát hiện đối tượng trong ảnh vệ tinh. Ngoài ra, bài báo cũng đề cập đến các ứng dụng thực tế của việc phát hiện đối tượng trong ảnh vệ tinh, bao gồm giám sát môi trường, quản lý tài nguyên và ứng phó với thảm họa. Cuối cùng, bài báo đã đưa ra những hướng nghiên cứu tiềm năng trong tương lai, như phát triển các mô hình hiệu quả hơn, xử lý các đối tượng nhỏ và tận dụng các nguồn dữ liệu đa dạng.

Từ khóa: *ảnh vệ tinh, học sâu, phát hiện đối tượng, thị giác máy tính.*

1. INTRODUCTION

In recent years, the application of artificial intelligence in various fields has brought about significant breakthroughs. In particular, computer vision, with its ability to analyze and understand images, has become an effective tool in many practical applications. One of the core research problems of computer vision is object detection, and among them, object detection in satellite images is attracting increasing attention.

Satellite images provide a huge source of data about the Earth, with increasingly high resolution and detail. However, extracting useful information from these images requires complex algorithms and models (Li et al., 2019). Object detection in satellite images is a difficult problem, requiring solving challenges such as different resolutions, noise, diversity of objects, and changes in environmental conditions.

Solving this problem successfully will open up many important applications in areas such as environmental monitoring, urban management, agriculture, military, and disaster relief (Li et al., 2022; Wang et al., 2023). For example, detecting changes in forests, urbanization, or unusual events such as wildfires and floods can help us make more effective management decisions.

This paper provides a comprehensive overview of object detection in satellite images, encompassing a range of topics from fundamental concepts to contemporary methods. The research delves into the unique challenges and characteristics of satellite imagery, offering readers a deeper understanding of the problem's complexity. Furthermore, the paper conducts a detailed comparison of anchor-based and non-anchor-based object detection methods, enabling readers to make informed decisions regarding the most suitable approach for their specific needs. Finally, the paper presents valuable suggestions for future research directions, paving the way for advancements in this field.

2. BACKGROUND

2.1. Common challenges in Object Detection Problem

Object detection faces several common challenges, which include:

Variation in object size: Objects can vary greatly in size, shape, orientation, and appearance within an image, depending on the resolution, angle, and illumination of the satellite. Satellite images are often large, complex, and have many noisy objects, and require significant preprocessing to extract useful information (Xia et al., 2018).

Lack of labeled data: Object detection demands a large amount of data to train and evaluate detection models. However, data labeling is time-consuming and labor-intensive, requiring human attention and expertise. This is especially true for satellite images, where the objects of interest are often small, complex, and diverse (Wang et al., 2023).

Low-resolution images: In low-resolution images, small objects often appear as a few pixels or even sub-pixel entities. This lack of detail makes it difficult to distinguish the object from the surrounding background noise or other objects (Wang et al., 2023). Low-resolution images contain less information overall, limiting the features that can be extracted by object detection algorithms. This can significantly impact the accuracy of the detection process. (James & Randolph, 2011).

Multiple objects in the same image: Images containing many objects, especially objects of different sizes, increase the complexity of the detection problem (Wang et al., 2023).

Noise and lighting variations: Noise and lighting variations in images also affect object detection.

Processing speed: Real-time performance is a challenge since object detection in satellite images tends to be real-time, detection speed also poses a significant challenge to detection algorithms. Because of the physical limitations of the processor for space-based applications, the characteristics of satellite data (presented in section 2.2) make object detection a challenging task due to the lack of adequate

datasets to train the network, and processing large satellite images on limited devices requires resources that are not always available in space environments (Lofqvist & Jose, 2021).

Therefore, the data needs to be diverse, high-quality, and suitable for the specific object detection task. Another challenge is labeling data and drawing bounding boxes for objects in the image (Xia, et al., 2018). Labels need to be accurate and consistent across different images and datasets and follow a clear and standardized annotation protocol. Inaccurate or inconsistent labels can negatively affect the performance and reliability of the object detection system.

2.2. Satellite image characteristics

Object detection is a challenging task with satellite images because the fundamental characteristics of satellite images are very different from conventional images (Ye et al., 2020; Aleissaee et al., 2023). Specifically, satellite images are captured from a panoramic view and have a large image range with comprehensive information, unlike natural images captured by ground-based cameras with a horizontal view. The imbalance between the area of the detected object and the background, combined with the possibility of objects being easily confused with random features in the background, further increases the complexity (Ye et al., 2020; Cole & Czerkawski, 2021).

There are five types of resolution when discussing satellite imagery in remote sensing: spatial, spectral, temporal, radiometric, and geometric (James & Randolph, 2011).

Satellite photos are often taken at high spatial resolution (hundreds of megapixels), and objects in the photos will have large differences in size. For instance, aircraft, vehicles, and ships appear small in high-resolution photos (about 0.5m/pixel), while large objects such as airports, streets, or large buildings appear larger in medium-resolution photos (1m/pixel). Large objects are often easier to detect, while small objects are often obscured by background information and are therefore more difficult to detect.

The quality of images taken from satellites also varies widely. Photos with poor quality are difficult to use for object detection because they may be noisy or have overlapping objects. That is why people often use high-resolution images, such as 30cm RGB, for object detection in remote sensing (Cole & Czerkawski, 2021). The temporal resolution feature (James & Randolph, 2011) makes it possible to take pictures at different times of the day and different seasons to produce different photos.

2.3. Satellite image sources

Satellite images can be obtained from various sources, including commercial and government satellites. Some of the popular databases that provide satellite images are USGS Earth Explorer, LandViewer, Copernicus Open Access Hub, Sentinel Hub, NASA Earthdata Search, Remote Pixel, and INPE Image Catalog. Apart from these, there are also open-source satellite image databases such as Google Earth Pro or Bing Maps which are regularly updated. Table 1 presents some useful information about open-source satellite image databases that are commonly used for scientific research, while an example of images from Google Earth is shown in Figure 1.

Figure 1. An image captured from Google Earth

Table 1. Some popular databases for the problem of detecting objects in satellite images

Data set	Number of photos	Variant	Size photo	Object class	Year
NWPU VHR-10	800	3775	~1000	Airplanes, ships, tanks, baseball fields, tennis courts, basketball courts, dirt fields, ports, bridges and vehicles	2014
VEDAI	1210	3640	1024	Cars, pickup trucks, vans, airplanes, boats, campers, tractors, vans and more	2015
UCAS-AOD	910	6029	1280	Cars, trucks	2015
DLR-3K	20	14235	5616	Cars, trucks	2015
HRSC2016	1061	6965	~1000	Ship	2016
RSOD	976	6950	~1000	Airplanes, overpasses, playgrounds, oil tanks	2017
DOTA	2806	188282	800–4000	Baseball fields, basketball courts, bridges, ports, helicopters, ground stadiums, large vehicles, airplanes, ships, small cars, football fields, tanks, swimming pools, tennis courts rackets and roundabouts	2017
DIOR-R	23463	192472	800	Windmills, Vehicles, Railway stations, Tennis courts, Storage tanks, Ships, Harbors, Stadiums, Land courses, Golf courses, Highway toll stations, Highway service areas, Dams, Chimneys, Bridges, Overpasses, Basketball Courts, Baseball Fields, Airports, Airplanes.	2022
EAGLE	8280	215986	936	Small vehicles (cars, trucks, transport vehicles, SUV, ambulances, police cars), large vehicles (trucks, large trucks, minibuses, buses, fire trucks, construction vehicles, trailers).	2020
GF1-LRSD	4406	7172	512	Ship	2021
SADD	2966	7835	224	Plane	2022

2.4. Performance indicators of object detection

In this section, we will discuss the most commonly used methods for evaluating the performance of object detection algorithms. These methods include Intersection over Union (IoU), precision, accuracy, recall, average precision (AP), and mean average precision (mAP) (Wang et al., 2023).

Intersection over Union (IoU) is a measure of the overlap between two bounding boxes – the predicted box and the actual box (Wang et al., 2023). When an object is detected in an image, a bounding box is created. The IoU index indicates how similar the predicted label is to the actual label. The higher the IoU, the greater the intersection, and the smaller the union. In other words, the model has high accuracy when the IoU index is high. The IoU measure can be calculated as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{A \cap B}{A \cup B} \quad (1)$$

Precision is the ratio of correct predictions (matching the actual box) to the total number of predictions, so:

$$Precision = \frac{TP}{TP+FP} = \frac{Relevant\ retrieved\ instances}{All\ retrieved\ instances} \quad (2)$$

Recall sensitivity represents the number of correct predictions over the total number of actual boxes. This is an important indicator that shows whether the model found all the labeled samples in the image or not.

$$Recall = \frac{TP}{TP+FN} = \frac{Relevant\ retrieved\ instances}{All\ relevent\ instances} \quad (3)$$

The higher the average AP accuracy, the better the system's detection performance for a given type of object in the data set. From the precision and recall found above, we can draw a precision curve according to recall (PR curve) for each separate class. The average accuracy AP is the area under this PR curve.

The average mAP measure is the average index of the average accuracy of the object classes detected by the system. Higher mAP values indicate better detection performance for the entire dataset. The mAP value is calculated as follows:

$$mAP = \frac{1}{n} \sum_{k=1}^{n} AP_k \quad (4)$$

In which, AP_k is the average AP value of object k and n is the total number of object classes.

3. APPLICATION OF OBJECT DETECTION PROBLEM IN SATELLITE IMAGES BACKGROUND

3.1. Common challenges in Object Detection Problem

Deep learning is a branch of computer vision that applies artificial neural networks to solve various image-processing tasks. One of these tasks is object detection, which aims to locate and identify objects of interest in an image. Object detection problems in satellite images are similar to those in natural images, but they also have some specific challenges. For example, satellite images often have low resolution, high noise, and complex backgrounds. Moreover, satellite images can be used for many different purposes, such as monitoring land use, detecting changes, identifying crops, and assessing natural disasters. Therefore, object detection in satellite images requires not only image classification and segmentation but also regression and other techniques to handle these issues (Li et al., 2022).

One of the main applications of remote sensing data is image classification, which aims to assign meaningful categories to each image based on its content. For example, an image can be classified as "urban," "forest," "agricultural land," or "buildings" (such as stadiums, bridges, airports, parking lots). This type of classification is called image-level classification (Cole & Czerkawski, 2021). However, some images may contain multiple categories, such as a forest with a river or a city with mixed land use. In these cases, image-level classification may not be sufficient to capture the diversity and complexity of the image.

Image segmentation is a key technique in image analysis and computer vision (Cole & Czerkawski, 2021). It aims to partition an image into segments or regions that have semantic meaning. The image segmentation technique assigns a class label to each pixel in the image, effectively transforming the image from a 2D-pixel grid to a 2D-pixel grid with assigned class labels. One common use of image segmentation is road or building segmentation, where the objective is to detect and separate roads and buildings from other elements in an image. The technology can also be applied to classify land use and crop types using satellite imagery and aerial photography.

One of the applications of remote sensing is to estimate continuous variables from images, such as wind speed, the height of trees, or soil moisture (Cole & Czerkawski, 2021). These variables can be useful for forecasting natural hazards such as storms, tsunamis, and volcanic eruptions. A common deep learning approach for this task is to use convolutional neural networks (CNN) to extract features from image data, and then use a fully connected neural network (FCNN) to perform regression. FCNN is trained to learn the mapping function from input images to target outputs, providing predictions for the continuous variables of interest.

Cloud detection in remote sensing images is a crucial task because clouds can block the visibility of the underlying land features. This affects the accuracy and efficiency of remote sensing analysis because the blocked areas cannot be correctly interpreted. Various methods have been proposed to detect clouds in remote sensing images. Deep learning methods use convolutional neural networks (CNN) to effectively detect clouds in remote sensing images (Cole & Czerkawski, 2021). These models are trained on large collections of remote sensing images, enabling them to learn and generalize about the distinctive features and patterns of clouds. The resulting cloud mask can be used to locate cloud pixels and eliminate them. This method enhances the accuracy of remote sensing analysis and provides a better view of the land, even in cloudy conditions.

Change detection is a vital component of remote sensing analysis. It allows the tracking of landscape or object changes over time. The technique can be applied to identify a range of changes, including land use change, urban development, coastal erosion, deforestation, or simply the displacement of equipment, airplanes, ships, factories, etc. Change detection can be performed on a pair of images taken at different times or by analyzing multiple images collected over some time (Cole & Czerkawski, 2021).

Crop classification in remote sensing is the process of assigning different crop types to pixels or regions in an image or a series of images. It helps to understand the spatial distribution and composition of crops in a given area, which can be useful for tracking crop development and detecting crop anomalies. Deep learning methods such as convolutional neural networks (CNN) can be applied to crop classification (Cole & Czerkawski, 2021). The best method depends on the characteristics and complexity of the data, the required accuracy, and the computational resources available. However, the quality and resolution of the input data, as well as the availability of labeled training data, are crucial factors for the performance of crop classification.

Remote sensing images are used in natural disaster response. Remote sensing images can provide valuable information about the extent and severity of damage caused by a disaster, such as an earthquake, a hurricane, or a landslide. Remote sensing images can help to locate and quantify the damage to buildings and infrastructure, identify the areas that are inaccessible or isolated, estimate the area affected by the disaster, and assess the risk of secondary hazards, such as flooding (Cole & Czerkawski, 2021). Moreover, remote sensing images can also assist in monitoring and managing other types of natural hazards, such as wildfires, droughts, or floods. Remote sensing images can enable the detection and tracking of fire fronts, the evaluation of vegetation conditions, or the prediction and warning of flood events in real-time.

4. DEEP LEARNING METHODS FOR OBJECT DETECTION IN SATELLITE IMAGES

4.1. Anchor-based object detection method

Object detection is a challenging task that requires locating and identifying objects of various sizes and shapes in an image. One of the common techniques for object detection is to use anchor boxes, which are predefined rectangular regions that cover different scales and aspect ratios of the objects. Anchor boxes are overlaid on the image and serve as reference points for generating candidate bounding boxes that contain the objects (The MathWorks, 2023).

There are two main types of methods that use anchor boxes for object detection: two-stage methods and one-stage methods (Liu et al., 2019). Two-stage methods first generate a set of region proposals using a separate algorithm, such as selective search or region proposal network, and then refine and classify them using a convolutional neural network. One-stage methods directly predict the bounding boxes and their classes from the image using a single network, without relying on external region proposals. Two-stage methods tend to have higher accuracy but lower speed, while one-stage methods are faster but less accurate.

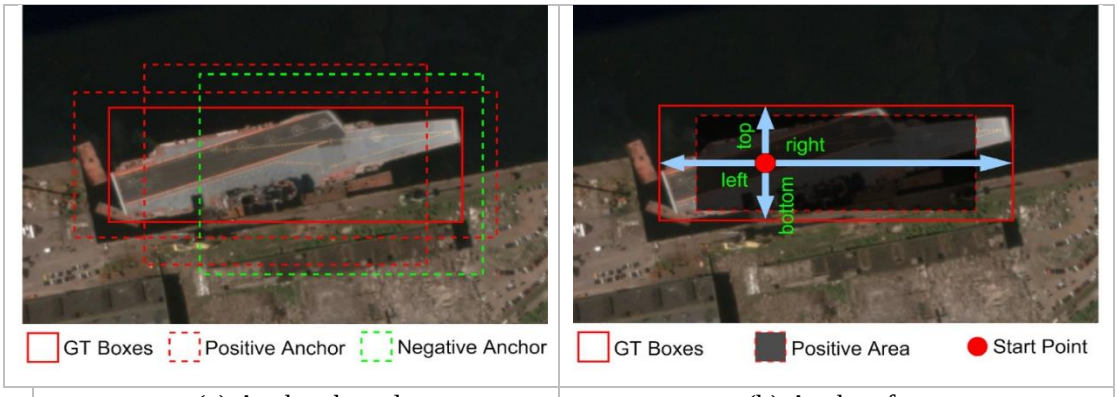
Anchor-based methods have some advantages, such as being able to detect multiple objects with different sizes and shapes, and handling occluded or overlapping objects. However, they also have some drawbacks, such

as requiring a large number of anchor boxes to cover the object space, which leads to an imbalance between positive and negative samples during training. Moreover, anchor boxes are fixed and cannot adapt to the shape of the objects, which makes it difficult to detect very large or small objects (Liu et al., 2020). Furthermore, anchor-based methods are computationally expensive and time-consuming to train and test.

Some of the popular models that use anchor-based methods are R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN, R-FCN, Mask R-CNN for two-stage methods, and Yolo, SSD, DSSD, RetinaNet, GA-RPN, M2Det for one-stage methods (Liu et al., 2020).

4.2. Anchor-free object detection method

Anchor-based methods use predefined anchor boxes that cover different scales and aspect ratios of objects. They assign labels to these boxes based on the overlap with the ground truth boxes. However, these methods have some drawbacks, such as the need to tune the anchor parameters and the difficulty of handling objects with extreme shapes. Anchor-free methods do not rely on predefined boxes, but rather use other ways to locate objects (Figure 2) (Wang et al., 2023).



(a) Anchor-based (b) Anchor-free

Figure 2. A visual explanation shows the difference between anchor-based and anchor-free methods. (a) The anchor-based methods predict the offsets based on a predefined anchor. (b) The anchor-free methods directly estimate the offsets of a point to its outside boundaries (Jin et al., 2024)

There are two subtypes of anchor-free methods: key point-based and centroid-based. Key point-based methods detect specific points on the objects, such as center points or corner points, and then group them to form bounding boxes (Liu et al., 2020). These methods can handle objects with arbitrary shapes, but they also have some limitations, such as the high computational cost of the grouping process and the low recall rate of the key points. Some examples of key point-based methods are CornerNet, ExtremeNet, CenterNet, and CentripetalNet (Cheng et al., 2018). Centroid-based methods use regions of interest or central locations of the objects to identify positive samples, similar to anchor-based methods. However, they do not need to generate anchor boxes, but rather predict the offsets and sizes of the bounding boxes from the regions or locations. These methods are simpler and faster than key point-based methods, but they may suffer from low accuracy for small objects or objects with large aspect ratios. Some examples of centroid-based methods are FoveaBox, FCOS, FSAF, and ObjectBox (Cheng et al., 2018).

4.3. Object detection method based on Transformer architecture

The object detection method based on the Transformer architecture is a neural network that uses a self-attention mechanism to capture dependencies between inputs. It consists of an encoder and a decoder with several transformer blocks, each containing a Multi-head Attention layer and a relay network (Figure 3) (Vaswani et al., 2017).

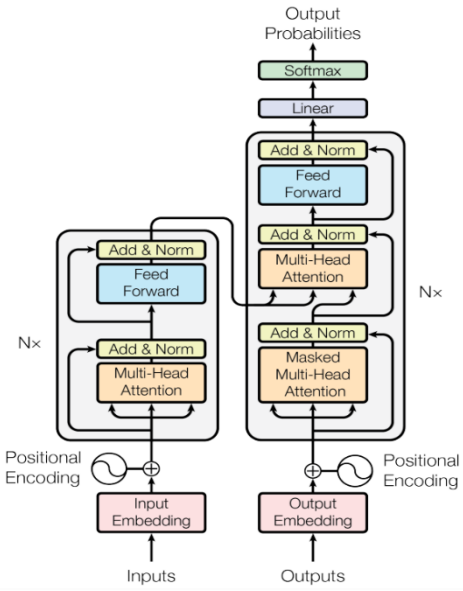


Figure 3. Transformer Architecture (Vaswani et al., 2017)

In recent years, the Transformer architecture has undergone significant advancements, especially in the field of Natural Language Processing (NLP) with the introduction of models like BERT and GPT-3. Many researchers are now interested in using this method in computer vision, given the impressive performance of the visual transformer (ViT) (Dosovitskiy et al., 2021).

The ViT is a visual model that is based on the original Transformer architecture and uses a self-attention mechanism to capture interactions between different components of an image. It does this by learning the relationships between these components. The ViT divides the image into fixed-sized arrays, which are then encoded and passed sequentially to the Transformer encoder. Several new variants of ViT have also achieved great success, including DeiT, PVT, TNT, and Swin (Figure 4). More recently, models like DALLE2/StableDiffusion and GPT-4 have also garnered significant attention (Han et al., 2022).

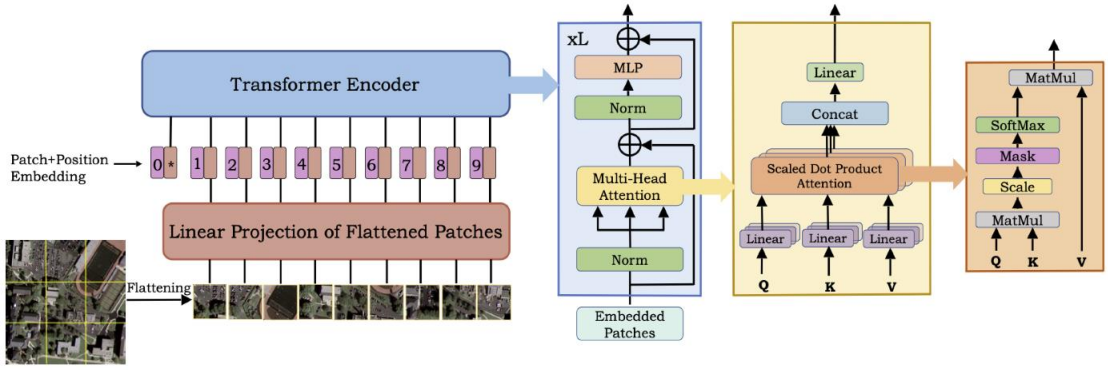


Figure 4. Architecture of ViT (Dosovitskiy et al., 2021)

5. CONCLUSION

Object detection in satellite images is a hot topic in the research field nowadays. We have introduced some deep learning techniques that achieve high performance in this task. We have also explored various applications of object detection in satellite imagery, such as geographic information systems, environmental monitoring, agriculture, and national security.

However, object detection faces some common difficulties, such as size variation, low resolution, lack of labeled data, and multiple objects in an image. The finding of this study is a valuable reference for researchers to enhance and create different deep learning models to improve the capability to detect characteristic objects in satellite images.

REFERENCES

Aleissaee, A. A., Kumar, A., Anwer, R. M., Khan, S., Cholakkal, H., Xia, G. S., & Khan, F. S. (2023). Transformers in Remote Sensing: A Survey. *Remote Sensing*. DOI: 10.3390/rs15071860

Cheng, G., Yang, C., Yao, X., Guo, L., & Han, J. (2018). When deep learning meets metric learning: remote sensing image scene

classification via learning discriminative CNNs. *IEEE Trans Geosci Remote Sensing*. DOI: 10.1109/TGRS.2017.2783902

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T.,… & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR 2021*. DOI: 10.48550/arXiv.2010.11929

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z.,… Xu, Y. (2022). A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence.* DOI: 10.1109/TPAMI.2022.3152247

James, C., & Randolph, W. (2011). *Introduction to Remote Sensing*. New York London: The Guilford Press.

Jin, C., Zheng, A., Wu, Z., & Tong, C. (2024). Transformer-Based Multi-layer Feature Aggregation and Rotated Anchor Matching for Oriented Object Detection in Remote Sensing Images. *Arabian Journal for Science and Engineering*. DOI: 10.1007/s13369-024-08892-z

Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., & Chanussote, J. (2022). Deep Learning in Multimodal Remote Sensing Data Fusion: A Comprehensive Review. *International Journal of Applied Earth Observation and Geoinformation*. DOI: 10.1016/j.jag.2022.102926

Li, W., Liu, H., Wang, Y., Li, Z., Jia, Y., & Gui, G. (2019). Deep Learning-Based Classification Methods for Remote Sensing Images in Urban Built-Up Areas. *IEEE Access, 7*, 36274-36284. DOI: 10.1109/ACCESS.2019.2903127

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2019). Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision, 128*, 261–318.

Liu, S., Zhou, H., Li, C., & Wang, S. (2020). Analysis of Anchor-Based and Anchor-Free Object Detection Methods Based on Deep Learning. *Proceedings of 2020 IEEE International Conference on Mechatronics and Automation.* DOI: 10.1109/ICMA49215.2020.9233610.

Lofqvist, M., & Jose, C. (2021). Optimizing Data Processing in Space for Object Detection in Satellite Imagery. *SmallSat 2021 – The 35th Annual Small Satellite Conference*. DOI: 10.48550/arXiv.2107.03774

Robin Cole & Mikolaj Czerkawski (2021). *Techniques for deep learning with satellite & aerial imagery*. Retrieved October 10, 2023, from https://github.com/satellite-image-deep-learning/techniques#4-object-detection

The Math Works. (2023). *Anchor Boxes for Object Detection*. Retrieved October 11, 2023, from https://www.mathworks.com/help/vision/ug/anchor-boxes-for-object-detection.html

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I,. (2017). Attention is all you need. *31st Conference on Neural Information Processing Systems.* DOI: 10.48550/arXiv.1706.03762

Wang, X., Wang, A., Yi, J., Song, Y., & Chehri, A. (2023). Small Object Detection Based on Deep Learning for Remote Sensing: A Comprehensive Review. *Remote Sensing.* DOI: 10.3390/rs15133265

Xia, G., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2018). DOTA: A large-scale dataset for object detection in aerial images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2018.00418

Ye, X., Xiong, F., Lu, J., Zhou, J., & Qian, Y. (2020). $\mathscr{F}3$-Net: Feature Fusion and Filtration Network for Object Detection in Optical Remote Sensing Images. *Remote Sensing*. DOI: 10.3390/rs12244027