Enhancing Electrical Grid Mapping with YOLO-NAS: A Deep Learning Approach to Satellite Imagery Analysis

By John Cline
Geog 870
Capstone

Abstract

This capstone paper explores the application of YOLO-NAS (You Only Look Once - Neural Architecture Search), a modern deep learning algorithm, in improving the mapping accuracy of electrical grid assets through satellite imagery analysis. Electrical grid mapping, crucial for efficient asset management and disaster recovery, currently relies heavily on manual inspection, which is prone to errors. YOLO-NAS, known for its real-time processing capabilities and high precision, is employed to automate the detection and localization of grid components such as wooden poles and streetlights. By leveraging YOLO-NAS, this study aims to reduce human error and increase the efficiency of grid mapping. The project utilizes a dataset of 2082 satellite images, annotated to identify these objects. The results demonstrate a mean average precision (mAP) of 54.7%, with precision and recall metrics indicating moderate success in object detection. While the performance of the model is promising, several limitations to the experiment exist, therefore, more work needs to be done to accurately assess the efficacy of the model to identify these objects.

Introduction

In the realm of Geographic Information Systems (GIS), organizations in the energy sector often face challenges in accurately locating and mapping electrical grid assets in both urban and rural areas. Tasks such as identifying the positions of assets like electrical poles from satellite imagery rely heavily on visual inspection. This can lead to positional errors in mapping, hindering planning, operations, and disaster recovery. Deep learning algorithms optimized for object detection offer a valuable solution to improve the efficiency and accuracy of mapping electrical grid assets. These algorithms can automate the detection and localization of grid components from satellite imagery, minimizing human error and variability. In this project, a modern computer vision deep learning algorithm called YOLO-NAS (You Only Look Once - Neural Architecture Search), known for its speed and accuracy, will be trained using a dataset of satellite imagery with wooden poles and streetlights labeled using common annotating techniques. The central question is: can modern object detection algorithms enhance electrical grid asset management?

Background Research

Deep Learning and Computer Vision

Deep learning is a subset of Artificial Intelligence (AI) and Machine Learning (ML). It was introduced in the early 2000s after other neural networks became popular. Early on, deep learning was not adopted due to its lack of scalability because of its demand for computing power and the lack of huge datasets. Since then, it has become a popular method for data analysis because of the growth of viable high-end computational resources and the abundance of available data (Roboflow 2020). Deep learning has seen major success in industries ranging from weather forecasting (Zaytar and Amrani 2016), stock market prediction (Rather et al. 2015), speech recognition (Sak et al. 2015), object detection (Liang 2015), character recognition (Zhang et al. 2017), landslide detection (Mezaal et al. 2017), time series prediction (Che et al. 2018), video processing (Xu et al. 2018), and many more.

Computer vision is a method of deep learning that focuses on enabling computers to identify and understand objects and people in images and videos (Khan et al. 2021). In a foundational sense, computer vision instructs machines to understand, grasp, and analyze high-level understanding of visual concepts – basically seeking to analyze images in the same way people do (Khan et al. 2021). Uses of computer vision include object recognition, object detection, video tracking, object segmentation, pose and motion estimation, scene modeling, and image restoration (Morris 2004). In this paper, I focus on the object detection method of computer vision that modern deep learning models employ.

Deep Learning Models and Remote Sensing

Deep learning has found various applications in the field of remotely sensed data, particularly in object detection. Beyond this, there are other key tools and techniques in computer vision that are worth noting. Image classification is one such technique, involving training models on diverse landcover types. This enables the models to accurately represent soil moisture, vegetation density, and topography (Affonso et al., 2017). Another significant application is change-detection, which quantitatively analyzes surface changes over time (Khelifi et al. 2022). This technique also leverages deep learning to process labeled landcover types and track changes in a specified area over time. The convergence of deep learning with satellite imagery in object detection has become a pivotal theme, particularly in the context of enhancing infrastructure mapping and grid efficiency. This is explored in various studies, including those by Groener et al. (2019), Nguyen et al. (2020), Huyan et al. (2021), Van Etten (2018), Gudžius et al. (2021), and Khan et al. (2021). These studies show how deep learning models optimized for object detection can identify small objects in satellite imagery.

YOLO (You Only Look Once)

Deep learning techniques related to remote sensing, including image classification, change detection, and object detection, rely on several foundational models. These include Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM), Restricted Boltzmann Machines (RBM), and Stacked Autoencoders (Voulodimos et al. 2018). CNNs are particularly prominent due to their effectiveness in representing spatial patterns and extracting features from remotely sensed data (Diwan et al., 2022). In the subfield of object detection, CNN-based models like YOLO (You Only Look Once) have proven especially proficient. YOLO is a single-stage model optimized for real-time object detection, demonstrating high accuracy and efficiency in identifying small objects in satellite imagery (Redmon 2016; Nguyen 2020). The model processes images in one pass, predicting bounding boxes and class probabilities simultaneously, achieving double the mean average precision (mAP) compared to other real-time detectors (Redmon 2016). The development of YOLO has continued with iterations up to YOLOv9, and now YOLO-World. YOLO-World introduces an AI version trained on various objects and enables users to prompt the model to identify specific items, building on the foundation laid by YOLOv9 (Roboflow 2023). These iterations address previous model limitations while retaining the core functionality of training and testing, advancing the field of object detection (Roboflow 2023).

Yolo-V8 and Yolo-NAS

YOLO-V8 is an advanced version of the YOLO object detection framework known for its improvements in speed and accuracy. Building on V8’s foundation, YOLO-NAS integrates Neural Architecture Search (NAS) to optimize the model's architecture (Medaramatla et al. 2024). This integration automates the process of finding the most efficient architecture, improving both performance and efficiency (Terven et al. 2023). YOLO-NAS also utilizes quantization-aware training blocks, maintaining high accuracy even when quantized to lower precision (Terven et al. 2023). This makes the model suitable for edge devices with limited computational resources (Terven et al. 2023). Moreover, AutoNAC (Automated Neural Architecture and Configuration) fine-tunes configuration parameters such as batch size and learning rate, ensuring the model is not only architecturally optimal but also configured to deliver the best performance for specific tasks (Terven et al. 2023; Saluky et al. 2024). This has led to significant improvements over previous YOLO versions in terms of precision and processing time, making YOLO-NAS an ideal choice for real-time applications, particularly for mobile and embedded systems (Roboflow 2023). I selected YOLO-NAS as the algorithm to test in this project because that is the model currently available in Roboflow's free tier. YOLO-NAS has been found appropriate to identify objects in various media such as surveillance footage (Saluky et al. 2024), detecting building boundaries (Tasyurek 2024), and smoke and wildfire detection in satellite imagery (Casas et al. 2023).

Model Evaluation

Common evaluation metrics for computer vision model performance include mean average precision (mAP), recall (R), and precision (P) (Harris and Glowacz 2021; Nguyen et al. 2020; Padilla et al. 2020; Padilla et al. 2021).

- **mAP (Mean Average Precision)**: This metric is the primary comparison point for all deep learning models and is largely based on IoU (Intersection over Union), which measures the overlap between the predicted and ground truth bounding boxes (Padilla et al. 2020).
- **IoU Probability Threshold**: Typically set at 0.5, this threshold determines whether a prediction is a true positive or a false positive (Nguyen et al., 2020; Padilla et al., 2020).
- **Precision (P)**: This measures the accuracy of positive predictions, indicating how many of the predicted bounding boxes match the ground truth (Padilla et al. 2020; Nguyen et al. 2020).
- **Recall (R)**: Recall gauges the model's ability to detect all relevant cases, reflecting how many of the ground truth bounding boxes were correctly predicted (Padilla et al. 2020; Nguyen et al. 2020).

Using these metrics provides an opportunity to objectively evaluate and compare model performance tested in this project.

Data and Methodology

Data acquisition

In developing a deep learning model capable of identifying electrical grid objects within satellite imagery, the construction of a comprehensive and well-prepared training dataset is paramount. This endeavor begins with the strategic acquisition of satellite images, which will serve as the foundation for training and subsequent analysis. I acquired 30cm resolution satellite images from Maxar's sample data which is freely available on their website (Maxar Technologies 2023). This dataset contains 2500 total images taken overhead of a certain region of Germany and is meant to be used to train deep learning algorithms on identifying solar panels. Even though the dataset has a different purpose, the images capture parts of the overhead electrical distribution network in the area.

Image Annotation

The process of building this training dataset involves meticulous preparation, including data annotation and image preprocessing. Annotation of datasets for deep learning applied to satellite and aerial imagery is a critical step in preparing for accurate object detection. Data annotation is a labor-intensive but critical step, requiring the manual delineation of electrical grid objects within the images. This process is facilitated by annotation tools such as Roboflow (Roboflow 2023). It streamlines the process of annotating images by providing tools for manual annotation. Roboflow supports a variety of annotation types, including bounding boxes, polygons, and segmentation masks, catering to different object detection and segmentation needs. Images were stored in Roboflow's cloud service and the electrical distribution objects were labeled using the bounding box method in the application (Roboflow 2023). I labeled two types of electrical distribution objects: wooden poles that are used to support overhead wires, and streetlights. These objects were categorized in the same class as "Pole". An example of this can be seen below (**Figure 1**). I annotated 824 images from the initial set and divided these images in the following manner: 629 training images, 111 validation images, and 84 test images.

Figure 2 *shows the process of labeling electrical distribution structures in the acquired satellite imagery.*

Image Preprocessing

Image preprocessing is a crucial step in training an algorithm, aimed at ensuring that datasets are standardized and ready for training models. Roboflow provides several preprocessing options. There are a few steps in the image preprocessing process that I have outlined here. The first step is auto-orientation. This step strips the EXIF metadata from images, ensuring they are displayed in the same orientation as they are stored on disk (Roboflow 2023). This prevents inconsistent orientations that might silently ruin object detection models. The next step is to resize each image. Resizing pixels modifies the image size to a specified set of dimensions. I set the image resizing to 640 x 640; this setting is recommended by Roboflow because it is optimal for model training (Roboflow 2023). The third step in preprocessing is adjusting the contrast of each image.

This step helps balance contrast across the dataset, making it easier for neural networks to detect edges and improve overall model performance (Roboflow 2023).

Data Augmentation

Data augmentation is a critical step in training robust computer vision models, particularly in the Roboflow platform, where it serves to enhance the training dataset and improve model performance. Data augmentation addresses the limitation that collected datasets might not capture every real-world scenario a model could encounter (Roboflow 2023). By creating variations from existing data, it provides new training samples, allowing the model to generalize better across a broader range of situations. Roboflow offers a range of augmentation techniques. I chose to flip each image creating a mirror image from each original (**Figure 3**) as well as to rotate each image along its horizontal and vertical axis (**Figure 3**) (Roboflow 2023). This extended my initial training dataset from 629 images to 2082 images.

Figure 4 *demonstrates how an image is flipped to create a mirror image of the original. This is a common computer vision data augmentation technique.*

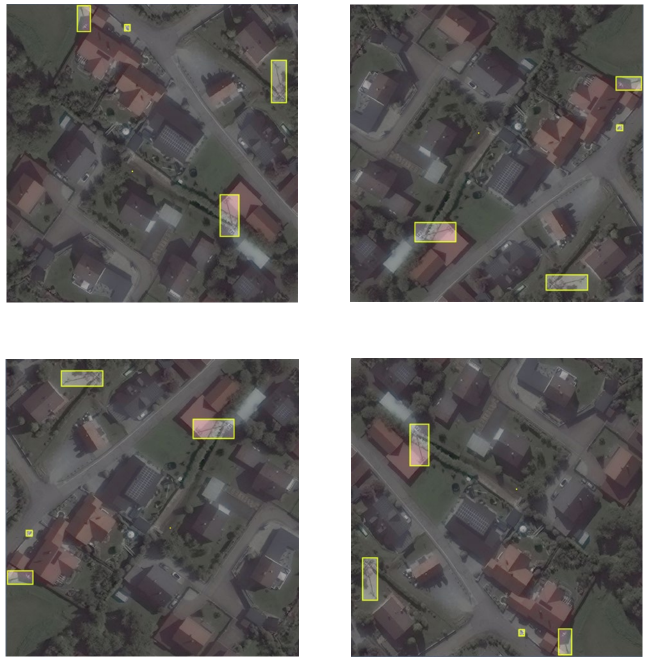

Figure 5 demonstrates rotating images along horizontal and vertical axes to create copies of the original image. Another common data augmentation technique.

Results

The algorithm was trained in Roboflow in three individual sessions using progressively more images, as more images were added to the dataset. The third and final session achieved the highest metrics discussed above: mean average precision (mAP) of 54.7%, 74.4% precision (P), and 40.6% recall (R).

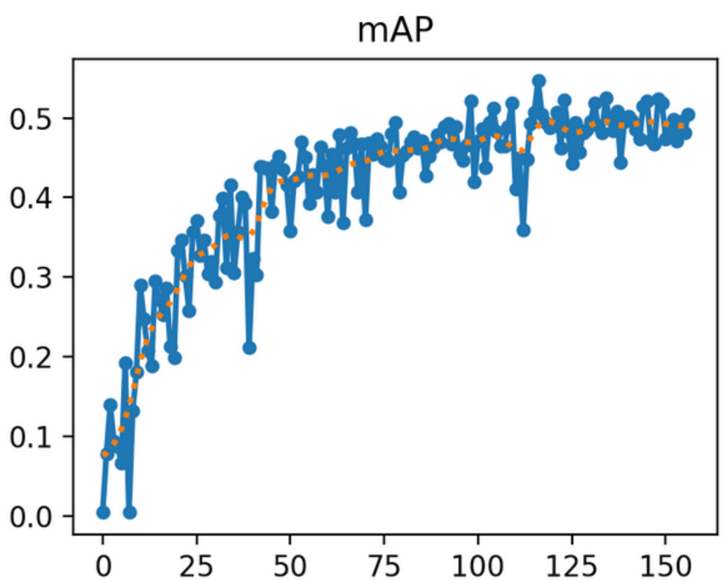

Figure 6 illustrates the mean average precision (mAP) metric during the final training session of the YOLO-NAS algorithm.

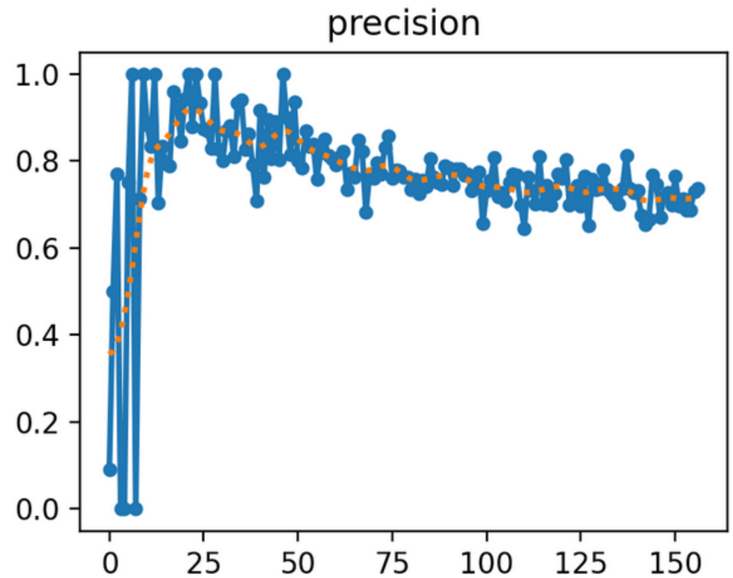

***Figure** 7 illustrates the precision (P) during the final training session of the YOLO-Nas algorithm.*

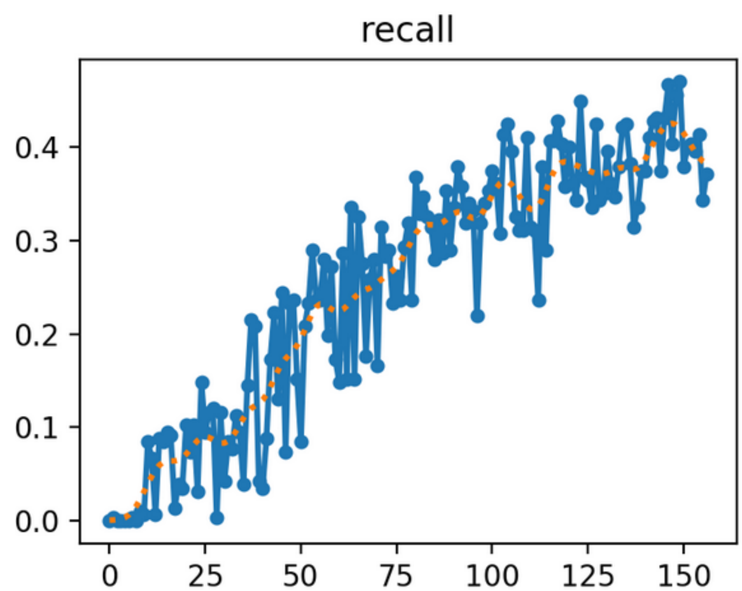

***Figure 8** illustrates the recall (r) of the algorithm during the final training session.*

***Figure 9** illustrates example images from the final training session. In the **left** image the model produced 5 true positives and one false positive, correctly predicting the presence of electric poles 5 of 6 times. In the **right** image, the model produced 3 false positives, incorrectly predicting the presence of electric poles 3 of 3 times.*

Analysis

The mAP result (Figure 6) shows how the mAP evolved throughout the training session with data points marking the mAP scores at specific iterations (Roboflow 2023). The upward trend indicates an improvement in the model's ability to make an accurate prediction over time and indicates that the algorithm is converging towards higher accuracy and then it plateaus toward the end suggesting that training has reached a point where further improvements are marginal and that the model has learned as much as possible.

Precision measures the accuracy of positive predictions by indicating the proportion of correctly identified objects (true positives) out of all predicted objects. The precision graph (**Figure 10**) illustrates a general upward trend where the model is increasingly accurate at identifying true positives to false positives, suggesting the model improves over time.

Recall (**Figure 8**) measures the model's ability to detect all relevant instances of an object, or how many of the ground truth bounding boxes were correctly predicted by the model. It evaluates the proportion of true positives among all actual positives. The graph's progression indicates that the model's precision improves over time, reflecting the model's increasing ability to correctly identify objects.

Discussion

The Mean Average Precision (mAP) of 54.7%, indicates that the model is moderately successful in accurately predicting bounding boxes and classifying them correctly. The graph also shows an upward trend, with convergence indicating that further improvements become marginal over time. The precision score of 74.4% suggests that a significant proportion of the model's positive predictions are accurate, matching ground truth bounding boxes. However, fluctuations in the graph may indicate variability in model performance across different test data. The recall score of 40.6% indicates that the model successfully detects relevant cases for a substantial portion of ground truth instances. The graph shows an upward trend with some variability, suggesting potential for further training to stabilize its detection capabilities.

The convergence patterns in the precision and recall graphs demonstrate the model's progress towards a balanced performance. Precision and recall scores reflect a trade-off: while the model is accurate in its positive predictions, it struggles to detect all relevant cases. The mAP score consolidates these measures, indicating an overall moderate performance. Further training iterations and dataset tuning could refine the model's accuracy and detection capabilities.

It is my presumption that these results reflect considerations that were made during the annotation process. The algorithm itself, YOLO-NAS, has been found to be successful in recent research at identifying a number of different objects (Saluky et al 2024; Casas et al. 2023; Tasyurek 2024; Anand et al. 2023). There is little work done related to YOLO-NAS's efficacy to identify smaller objects in satellite imagery such as electric utility poles. However, since NAS is simply V8 with optimized neural architecture there is little evidence here to suggest the model is incapable of identifying these objects. Likely, the consideration to classify all objects together (wooden poles with cross arms and streetlights) "confused" the model. This is suggested in the fluctuations in precision and recall. During the annotation process, I found there to be far fewer large wooden electric poles used for structure of overhead wires. The ratio of wooden poles to streetlights in the imagery was almost 20:1 in favor of streetlights. Similarly, the consideration of capturing the shadows of these objects could have affected this result. It was difficult to see either type of pole in the image unless the object casted a shadow in the image, this likely led to specific occurrences of false positive where the model predicted an object was present but there was no object in the image such as the example in **Figure 9 Left**. In these cases, straight lines like roadside curbs or fences look like shadows of poles.

Future Work

It is evident the annotation process needs to be refined. Classifying the types of poles separately will likely strengthen model performance. Furthermore, while Roboflow's ease of use was high, going through this process on open-source software would likely increase overall control of each step and result in a more robust and refined model. More work needs to be done to assess the initial research question.

References

Affonso, C., Rossi, A. L. D., Vieira, F. H. A., & de Leon Ferreira, A. C. P. (2017). Deep learning for biological image classification. *Expert systems with applications*, *85*, 114-122.

Anand, R. N., Singh, R. P., Gupta, D., & Palaniappan, K. (2023, December). Ship Detection in Satellite Images using You Only Look Once-Neural Architecture Search. In *2023 9th International Conference on Signal Processing and Communication (ICSC)* (pp. 463-468). IEEE.

Casas, E., Ramos, L., Bendek, E., & Rivas-Echeverría, F. (2023). Assessing the effectiveness of YOLO architectures for smoke and wildfire detection. *IEEE Access*.

Che Z, Purushotham S, Cho K, Sontag D, Liu Y (2018) Recurrent neural networks for multivariate time series with missing values. Sci Rep 8(1):1–12.

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, *82*(6), 9243-9275.

Groener, A., Chern, G., & Pritt, M. (2019, October). A comparison of deep learning object detection models for satellite imagery. In *2019 Intsitute of Electrical and Electronics Engineers: Applied Imagery Pattern Recognition Workshop (AIPR)* (pp. 1-10). IEEE.

Gudžius, P., Kurasova, O., Darulis, V., & Filatovas, E. (2021). Deep learning-based object recognition in multispectral satellite imagery for real-time applications. *Machine Vision and Applications*, *32*(4), 98.

Haris, M., & Glowacz, A. (2021). Road object detection: A comparative study of deep learning-based algorithms. *Electronics*, *10*(16), 1932.

Huyan, L., Bai, Y., Li, Y., Jiang, D., Zhang, Y., Zhou, Q., ... & Cui, T. (2021). A lightweight object detection framework for remote sensing images. *Remote Sensing*, *13*(4), 683.

Khan, A. A., Laghari, A. A., & Awan, S. A. (2021). Machine learning in computer vision: a review. *EAI Endorsed Transactions on Scalable Information Systems*, *8*(32), e4-e4.

Khan, S. D., Alarabi, L., & Basalamah, S. (2022). A unified deep learning framework of multi-scale detectors for geo-spatial object detection in high-resolution satellite images. *Arabian Journal for Science and Engineering*, 1-16.

Khelifi, L., & Mignotte, M. (2020). Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *Ieee Access*, *8*, 126385-126400.

Maxar Technologies. (2023). *15 cm HD and 30 cm View-Ready Solar Panels, Germany*. [Online]. Available at: https://resources.maxar.com/product-samples/15-cm-hd-and-30-cm-view-ready-solar-panels-germany [Accessed 12 March 2024].

Medaramatla, S. C., Samhitha, C. V., Pande, S. D., & Vinta, S. R. (2024). Detection of Hand Bone Fractures in X-ray Images using Hybrid YOLO NAS. *IEEE Access*.

Mezaal MR, Pradhan B, Sameen MI, Shafri M, Zulhaidi H, Yusoff ZM (2017) Optimized neural architecture for automatic landslide detection from high resolution airborne laser scanning data. Appl Sci 7(7):730. https://doi.org/10.3390/app7070730.

Morris T (2004) Computer Vision and Image Processing, Palgrave Macmillan Ltd, 1st edition, pp 1–320.

Nguyen, N. D., Do, T., Ngo, T. D., & Le, D. D. (2020). An evaluation of deep learning methods for small object detection. *Journal of electrical and computer engineering*, *2020*, 1-18.

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237-242). IEEE.

Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., & Da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, *10*(3), 279.

QSJXYZ. (2020). Electric Pylon Detection in RSI. GitHub. Retrieved Date February 8, 2024, from https://github.com/qsjxyz/Electric-Pylon-Detection-in-RSI.

Rather AM, Agarwal A, Sastry VN (2015) Recurrent neural network and a hybrid model for prediction of stock returns. Expert Syst Appl 42(6):3234–3241.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Roboflow. (2020). Why Preprocess and Augment?. Roboflow Blog. Available at: https://blog.roboflow.com/why-preprocess-augment/ [Accessed April 30, 2024].

Roboflow. (n.d.). How to Use Roboflow with Aerial Imagery. Roboflow Blog. Retrieved [Date you accessed the article], from https://blog.roboflow.com/how-to-use-roboflow-with-aerial-imagery/.

Roboflow. (2023). *Roboflow Documentation*. Available at: https://docs.roboflow.com [Accessed April 30, 2024].

Saluky, S., Nugraha, G. B., & Supangkat, S. H. (2024). Enhancing Abandoned Object Detection with Dual Background Models and Yolo-NAS. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(2), 547-554.

Sak H, Senior A, Rao K, Beaufays F (2015) Fast and accurate recurrent neural network acoustic models for speech recognition. arXiv preprint arXiv:1507.06947.

Satellite Image Deep Learning. (n.d.). Techniques for deep learning with satellite & aerial imagery. GitHub. Retrieved February 8, 2024, from https://github.com/satellite-image-deep-learning/techniques.

Satellite Image Deep Learning. (n.d.). Annotation of datasets for deep learning applied to satellite and aerial imagery. GitHub. Retrieved February 8, 2024, from https://github.com/satellite-image-deep-learning/annotation.

Tasyurek, M. (2024). BBD: a new hybrid method for geospatial building boundary detection from huge size satellite imagery. *Multimedia Tools and Applications*, 1-31.

Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, *5*(4), 1680-1716.

Van Etten, A. (2018). You only look twice: Rapid multi-scale object detection in satellite imagery. *arXiv preprint arXiv:1805.09512*.

Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018:1–13.

Qiao, S., Sun, Y., & Zhang, H. (2020). Deep learning based electric pylon detection in remote sensing images. *Remote Sensing*, *12*(11), 1857.

Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, *5*(4), 1680-1716.

Xu N, Liu AA, Wong Y, Zhang Y, Nie W, Su Y, Kankanhalli M (2018) Dual-stream recurrent neural network for video captioning. IEEE Trans Circuits Syst Vid Technol 29(8):2482–2493. https://doi.org/10.1109/TCSVT.2018.2867286.

Zaytar MA, El Amrani C (2016) Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Int J Comput Appl 143(11):7–11.

Zhang, W., Witharana, C., Li, W., Zhang, C., Li, X., & Parent, J. (2018). Using deep learning to identify utility poles with crossarms and estimate their locations from google street view images. *Sensors*, *18*(8), 2484.

Zhang XY, Yin F, Zhang YM, Liu CL, Bengio Y (2017) Drawing and recognizing chinese characters with recurrent neural network. IEEE Trans Pattern Anal Mach Intell 40(4):849–862.