Evaluating YOLO models in Power Lines Detection using TTPLA Dataset

Hamza Kaddour[†§1] and Mostafa M. Fouda[*†§2]
[*]Department of Computer Science, Idaho State University, Pocatello, ID 83209, USA.
[†]Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA.
[§]Center for Advanced Energy Studies (CAES), Idaho Falls, ID 83401, USA.
Emails: [1]hamzakaddour@isu.edu, [2]mfouda@ieee.org

***Abstract*—Automated inspection of power grid infrastructure, including wires and transmission towers, is essential for ensuring the safety and reliability of electricity distribution systems. This study evaluates the performance of state-of-the-art object detection algorithms from the YOLO (You Only Look Once) family (YOLOv5, YOLOv8, and YOLOv11) using the Transmission Tower and Power Line Aerial-Image (TTPLA) dataset. The TTPLA dataset, characterized by high-resolution aerial imagery with annotated bounding boxes, presents challenges such as complex backgrounds, diverse object geometries, and class imbalance. Among the models tested, YOLOv8 achieved the highest precision and recall metrics, demonstrating exceptional accuracy in detecting wires and transmission towers under real-world conditions. By leveraging advanced deep learning techniques, addressing class imbalance, and fine-tuning hyperparameters, this work establishes a new benchmark for UAV-based power grid inspections, highlighting the potential of cutting-edge computer vision models in infrastructure monitoring.**

***Index Terms*—Computer Vision, Wire, Transmission Tower, Object Detection, YOLO.**

I. Introduction

Efficient monitoring and maintenance of power grid infrastructure are critical for ensuring the safety and reliability of electricity distribution. Traditional methods of inspecting power lines and transmission towers, such as manual surveys, helicopter-assisted inspections, or crawling robots, are often time-consuming, expensive, and less adaptable to real-time scenarios. With the advent of unmanned aerial vehicles (UAVs) [1], there has been a paradigm shift toward automated power grid inspections. UAVs offer a cost-effective and flexible solution, enabling the collection of high-resolution aerial imagery over large areas in a short time. However, extracting actionable insights from these images demands advanced computer vision techniques capable of detecting and localizing power lines and transmission towers under challenging conditions.

Detecting power lines and transmission towers poses unique challenges due to the nature of the objects and their environments [2]. Power lines are long and thin and often blend into complex backgrounds like vegetation, urban structures, or open skies, making them difficult to distinguish. Transmission towers, on the other hand, exhibit diverse structures and shapes, with variations in materials, lighting conditions, and viewpoints further complicating the detection task. These challenges necessitate robust and accurate detection models that can operate effectively in real-world scenarios.

To address these challenges, researchers have developed datasets such as TTPLA [3], which provide annotated aerial images of power lines and transmission towers for training and evaluating detection models. However, the baseline performance reported on these datasets indicates significant room for improvement. For instance, the average precision (AP) for bounding boxes and instance masks in the TTPLA dataset remains relatively low, highlighting the need for more advanced methodologies.

In this study, we present a comprehensive evaluation of three versions of the YOLO object detection model—YOLOv5, YOLOv8, and YOLOv11—on a dataset containing power lines and transmission towers. YOLO models are renowned for their balance between speed and accuracy, making them ideal for real-time applications like UAV-based inspections. Our experiments demonstrate that the latest versions of YOLO significantly outperform baseline models in terms of average precision and other metrics. By leveraging advancements in neural network architectures and fine-tuning strategies, we achieve superior detection performance, setting a new benchmark for automated power grid inspections.

The remainder of this paper is organized as follows: Section 2 discusses related work on power line and transmission tower detection, including existing datasets and detection models. Section 3 describes the dataset and methodology used in our experiments. Section 4 presents the experimental results and analysis, and Section 5 concludes the paper with insights and future directions for research.

II. Literature Review

Accurate detection of power lines and transmission towers is critical for automated power grid inspection, with various methods explored over the years. Early approaches relied on traditional techniques such as edge detection and Hough Transform, which performed reasonably well in simple scenarios but struggled in cluttered backgrounds and under varying conditions [4], [5]. Template matching methods were also employed for transmission tower detection but suffered from limitations related to scalability and robustness [6].

The rise of deep learning has significantly advanced object detection in this domain. Single-stage detectors like YOLO (You Only Look Once) have gained prominence for their speed and efficiency. Variants such as YOLOv5, YOLOv8, and YOLOv11 have shown remarkable improvements in detecting objects of varying scales, making them ideal for real-time UAV-based applications [7]. SSD (Single Shot Detector) models have also been utilized, particularly for small objects like power lines, although they often require additional fine-tuning to achieve high precision [8].

Two-stage detectors, including Faster R-CNN and Mask R-CNN, have been explored for higher accuracy. While these models excel in crowded or complex scenes, their computational requirements make them less suitable for real-time UAV operations [9], [10]. Additionally, segmentation-based methods like Yolact have been applied to enhance instance-level understanding, though these often compromise speed for accuracy [10].

Several studies have introduced specialized models tailored to the unique challenges of detecting thin and long structures like power lines. Architectures such as Wire-Net focus on pixel-level accuracy and are particularly effective in segmenting elongated objects [11]. Multi-modal approaches, combining RGB and thermal imaging, have shown promise in improving detection under adverse conditions [12].

Despite these advancements, challenges such as class imbalance, complex backgrounds, and crowded objects persist. Emerging trends in transformer-based models and domain adaptation techniques hold potential for addressing these gaps, particularly in scenarios with diverse environmental conditions [13].

With this paper, we contribute to filling the research gap in the literature, by taking into account a variety of algorithms and applying them in a Power line detection using the TTPLA dataset context for general-purpose use, which can be later applicable to different datasets with different scalability or deployment characteristics.

III. TTPLA Dataset

The TTPLA (Transmission Tower and Power Line Aerial-Image) dataset plays a pivotal role in advancing the development of robust detection methods for power grid monitoring. Introduced as a publicly available dataset, TTPLA consists of 1,100 high-resolution aerial images captured by UAVs, featuring a wide variety of transmission towers and power lines. These images are annotated at the pixel level, providing instance segmentation labels for a total of 8,987 objects. Unlike previous datasets that focus on coarse-level semantic segmentation or image-level classification, TTPLA offers detailed instance segmentation, making it a valuable resource for detecting and analyzing individual objects in crowded or overlapping scenarios.

The images in TTPLA were collected under diverse conditions, including varying lighting, weather, and backgrounds, to reflect real-world complexities. The dataset encompasses multiple tower structures (e.g., lattice, tubular steel, wooden poles) and intricate power line configurations, addressing challenges such as class imbalance, scale variations, and occlusions. Notably, the dataset features a high degree of background complexity, with power lines often blending into urban or natural scenes, thereby presenting unique challenges for object detection models.

Baseline evaluations on TTPLA using state-of-the-art models like Yolact [14] with ResNet [15] backbones demonstrated average precision (AP) scores as low as 22.96% for bounding boxes and 15.72% for instance masks. These results highlight the difficulty of detecting thin and elongated structures like power lines and the need for advanced detection methods to achieve higher accuracy [3]. This study builds upon the foundation laid by TTPLA, utilizing improved detection models (YOLOv5, YOLOv8, YOLOv11) to set new benchmarks and address the challenges inherent in aerial image analysis.

IV. System Model

In this section, we present the dataset used and the system model on which we based our study.

A. Overview

The architecture of our solution is described in Figure 1 and is divided into four main steps: data acquisition (TTPLA), data preparation, labels creation, models training, and evaluation.

B. Data preparation

For the TTPLA dataset, the data preparation process involved meticulous refinement and organization to ensure high-quality inputs for training. First, any instances labeled as void were removed from the JSON annotation files to eliminate irrelevant or ambiguous data. The dataset was then bifurcated into separate folders for training, testing, and validation. Care was taken to ensure that the distribution of labels was representative, with most of the key labels appearing across all three subsets. The dataset's labels were mapped as follows: 'cable' (0), 'tower_lattice' (1), 'tower_tucohy' (2), and 'tower_wooden' (3). This organized structure provided a balanced foundation for model development and evaluation.

C. Labels Creation

To prepare the dataset for training with YOLO models, we converted the annotation labels from JSON format to the required TXT format [16]. The original JSON annotations contained detailed information about the bounding boxes, including the coordinates,

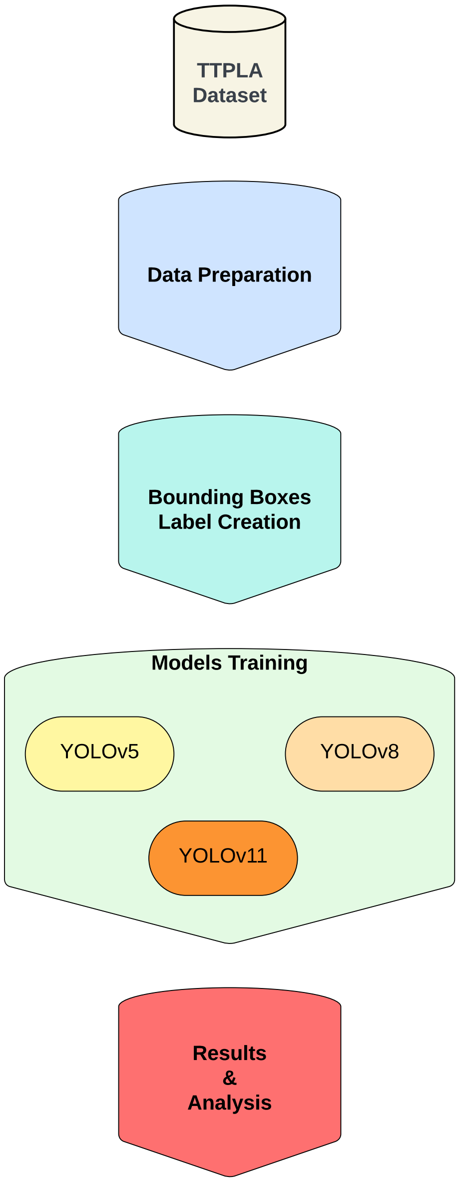

Fig. 1. System model architecture.

dimensions, and class labels for each object. A custom script was developed to parse these JSON files and extract the necessary bounding box parameters: normalized center coordinates (x, y), width, height, and the corresponding class indices. These values were then formatted into YOLO-compatible TXT files, where each line represents a bounding box with the structure: class, x_center, y_center, width, and height. This conversion ensured seamless integration of the dataset into the YOLO training pipeline, as the TXT format allows efficient processing by YOLO models while adhering to their input requirements.

D. Models specifications

In this phase, we applied the YOLOv5, 8, and 11 to the image dataset, and we performed a comparative analysis based on the performance metrics.

1) Algorithms used: YOLO (You Only Look Once) is a deep learning framework renowned for its efficiency and accuracy in object detection tasks. By processing the entire image in a single forward pass, YOLO achieves real-time detection capabilities, making it ideal for applications such as wire detection, where precision and speed are critical. The framework's evolution has led to multiple versions, each improving on aspects like accuracy, speed, and flexibility.

- **YOLOv5:** This version is lightweight and fast, making it highly suitable for real-time applications such as wire detection. It benefits from extensive community support and a wide range of pre-trained models for easy customization. However, its accuracy is slightly lower compared to newer versions, particularly for detecting thin and intricate objects like wires. It strikes a good balance between speed and performance, focusing on simplicity and ease of deployment.
- **YOLOv8:** Building on its predecessor, YOLOv8 features an improved model architecture and greater accuracy, making it better equipped to detect small or intricate objects such as wires in cluttered environments. It incorporates enhanced backbone networks and optimized training pipelines. Despite these improvements, it demands higher computational resources, which might limit its use on devices with constrained hardware capabilities.
- **YOLOv11:** As the latest iteration, YOLOv11 offers state-of-the-art performance with advanced detection capabilities for complex scenarios, such as detecting dense or overlapping wires. It integrates cutting-edge techniques like transformer-based architectures, improving its ability to capture fine details. However, it requires significant computational power and longer training times, and it has less community support compared to earlier versions, making it a less accessible option for some applications.

2) Evaluation Metrics:

a) Precision(B): Precision (B) is a metric that quantifies the proportion of correctly predicted bounding boxes out of all the bounding boxes predicted by the model. It measures the ability of the model to avoid false positives, ensuring that the detected objects are relevant and accurately classified. In the context of YOLO training for wire detection, high precision indicates that the model is effective at identifying wires and structures with minimal incorrect predictions. Precision is particularly critical when false detections can have severe consequences, such as in high-stakes monitoring scenarios. A higher precision reflects the model's confidence and accuracy in correctly classifying objects it predicts.

b) Accuracy(B): Accuracy (B) evaluates the overall correctness of the bounding box predictions by comparing both the localization and classification against ground truth labels. It combines elements of precision and recall to provide a holistic measure of the model's performance in detecting and localizing objects. In YOLO training, accuracy is essential for understanding how well the model generalizes across the dataset, capturing both positive and negative cases accurately. For wire detection tasks, accuracy ensures that the model identifies wires and towers while minimizing missed detections or incorrect classifications. This metric provides a comprehensive view of the model's performance across the dataset.

c) mAP50: Mean Average Precision at IoU 50% (mAP50) calculates the average precision across all classes for predictions with an Intersection over Union (IoU) threshold of 50%. This means a prediction is considered correct if its bounding box overlaps with the ground truth by at least 50%. In YOLO training, mAP50 serves as a foundational metric to evaluate the model's capability in detecting objects with sufficient spatial alignment. It is often used as an initial benchmark for performance, particularly in tasks like wire detection, where precise localization is necessary to distinguish between closely spaced objects.

d) mAP50-95: Mean Average Precision at IoU thresholds ranging from 50% to 95% (mAP50-95) is a more stringent metric that assesses the model's performance across varying levels of spatial alignment. This metric averages the precision values at multiple IoU thresholds in steps of 5% (e.g., 50%, 55%, ..., 95%), providing a comprehensive evaluation of both detection accuracy and localization quality. In YOLO training, mAP50-95 highlights the model's robustness in handling fine-grained object detections. For wire detection, this metric is especially valuable as it reveals the model's ability to handle intricate structures and overlapping objects.

e) Fitness: Fitness is a composite metric designed specifically for YOLO models, combining precision, recall, and other performance indicators into a single score. It provides an overall assessment of the model's suitability for the task at hand by weighing these metrics based on their importance. During YOLO training, the fitness score helps determine the optimal balance between detecting as many objects as possible and maintaining high confidence in predictions. In wire detection, fitness ensures the model achieves a practical trade-off, offering reliable and actionable detections without overcomplicating post-processing or decision-making.

V. Performance Analysis

In this section, we present the results of the analysis of the classification algorithms we applied to the TTPLA dataset.

Table I showcases the results of the implementations of the 3 YOLO versions on the dataset, and the calculations of the evaluation metrics.

For most of the evaluation metrics, the tested YOLO models outperformed the existing methods (Resnet-50 and Resnet-101).

Subsequently, the YOLOv8 outperformed its peers in most of the evaluation metrics for most of the aspect ratios, which gives solid performances in the mean average precision from 50 to 95 % and Fitness.

In addition, we carefully monitored the performance of the YOLOv8 model, and we plotted the loss functions and the evolution of the metrics for the 700x700 images in Figure 2. The plots show that the training process went smoothly, and the loss for most of the cases approached 0 relatively quickly.

VI. Conclusion

In this paper, We applied three YOLO versions (5, 8, and 11) on the TTPLA dataset to detect the power lines in the images. Our comparative study of the performance of the proposed models demonstrated that YOLOv8 showed the most solid overall performance for different aspect ratios (700x700, 550x550, and 640x360). Finally, the tested models showed better performance than the benchmark models applied in the dataset paper, resnet-50 and resnet-101.

Our future work will focus on including the other labels in our study and training to discern the different objects in the detection process and combining different object detection models in an ensemble model to get the best performance. Furthermore, we are projecting on working on transfer learning of the ensemble model on other wire and transmission towers images datasets and evaluating the results.

Acknowledgments

This research made use of the Idaho National Laboratory's High Performance Computing systems located at the Collaborative Computing Center and supported by the Office of Nuclear Energy of the U.S. Department of Energy and the Nuclear Science User Facilities under Contract No. DE-AC07-05ID14517.

References

[1] B. Cherif, H. Ghazzai, and A. Alsharoa, "Lidar from the sky: Uav integration and fusion techniques for advanced traffic monitoring," *IEEE Systems Journal*, vol. 18, no. 3, pp. 1639–1650, 2024.

[2] V. Gupta and V. Rathore, "Automatic detection and classification of transmission lines and towers in LiDAR data," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 286–290, 2018.

[3] R. Abdelfattah, X. Wang, and S. Wang, "TTPLA: An Aerial-Image Dataset for Detection and Segmentation of Transmission Towers and Power Lines," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.

[4] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[5] R. O. Duda and P. E. Hart, "Use of the hough transform to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

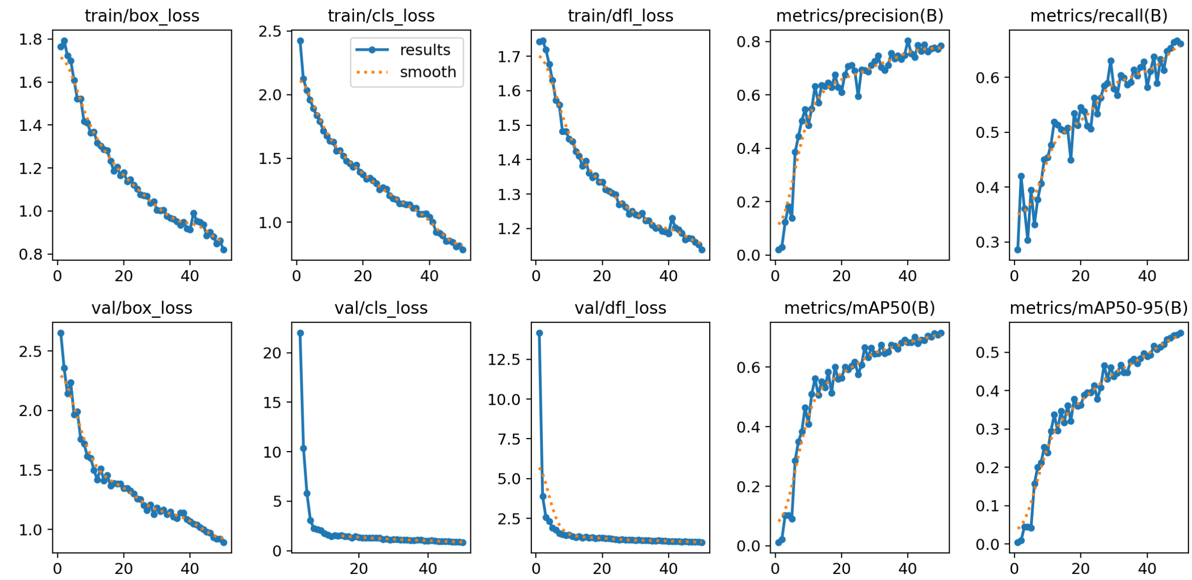

Fig. 2. Results of the training process of YOLOv8 on the 700x700 images.

TABLE I
PERFORMANCE EVALUATION OF THE YOLO MODELS COMPARED WITH THE EXISTING METHODS.

Model	Aspect Ratio	$mAP_{50\%}$	$mAP_{50-95\%}$	Precision (B)	Recall (B)	Fitness
ResNet-50	700x700	42.62	21.90	-	-	-
	550x550	43.37	20.76			
	640x360	**46.72**	16.50			
ResNet-101	700x700	43.19	22.96	-	-	-
	550x550	45.30	22.61			
	640x360	44.99	18.42			
YOLOv5	700x700	48.19	33.37	63.19	**48.51**	34.85
	550x550	45.34	30.79	59.13	46.63	32.25
	640x360	45.54	31.09	63.41	**44.49**	32.53
YOLOv8	700x700	**48.23**	**34.24**	**64.61**	48.26	**35.63**
	550x550	**45.93**	**31.66**	60.51	**47.33**	**33.09**
	640x360	45.45	**31.25**	**64.93**	44.43	**32.67**
YOLOv11	700x700	46.80	32.39	64.30	47.33	33.83
	550x550	44.93	30.91	**60.84**	45.86	32.31
	640x360	44.19	29.89	61.49	44.09	31.32

[6] R. C. Gonzalez and R. E. Woods, "Digital image processing," in *Proceedings of the IEEE*. Prentice Hall, 2002.

[7] G. e. a. Jocher, "Yolov5: Implementation of you only look once object detector," *GitHub Repository*, 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[8] W. e. a. Liu, "Ssd: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 21–37.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[10] K. e. a. He, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 2961–2969.

[11] O. e. a. Yetgin, "Power line recognition from aerial images with deep learning," *IEEE Transactions on Aerospace and Electronic Systems*, 2018.

[12] V. e. a. Nguyen, "Intelligent monitoring and inspection of power line components powered by uavs and deep learning," *IEEE Power and Energy Technology Systems Journal*, vol. 6, no. 1, pp. 11–21, 2019.

[13] N. e. a. Carion, "End-to-end object detection with transformers," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 213–229.

[14] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-Time Instance Segmentation," https://github.com/dbolya/yolact, 2019, accessed: 2024-12-10.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[16] Ultralytics, "JSON2YOLO: Convert JSON Annotations to YOLO Format," https://github.com/ultralytics/JSON2YOLO, 2023, accessed: 2024-12-10.