KAMPALA INTERNATIONAL UNIVERSITY

KIU
Journal of Science, Engineering and Technology

KJSET

Research Article

YOLOv9s-based surveillance of human and animal activities near optical fiber infrastructure

Enerst Edozie[1*], Aliyu Nuhu Shuaibu[1], Ukagwu Kelechi John[1], Bashir Olaniyi Sadiq[1]

[1] *Department of Electrical Engineering, Kampala International University, 20000 Ishaka, Uganda.*
edozie.enerst.11275@studwc.kiu.ac.ug[1*], nuhua@kiu.ac.ug[1], ukagwu.john@kiu.ac.ug[1], bosadiq@kiu.ac.ug[1]

Corresponding Author: Enerst Edozie edozie.enerst.11275@studwc.kiu.ac.ug

Paper history:
Received 05 March 2025
Accepted in revised form 08 May 2025

Keywords
Object Detection; YOLOv9s; Deep Learning; Real-Time Detection; Fiber Optic Infrastructure.

Abstract

This study presents an intelligent deep learning-based approach for anomaly detection in fiber optic infrastructure using the YOLOv9s object detection model. The objective is to improve surveillance efficiency by accurately detecting humans and animals interacting with pole-mounted fiber optic structures in real time. A custom dataset was developed by capturing and annotating real-world video footage from diverse locations and environmental conditions. High-resolution images were extracted and labeled using CVAT, ensuring high-quality annotations across object types and activities. Model training was conducted in PyTorch, incorporating data augmentation techniques such as brightness adjustment, flipping, noise injection, and geometric transformations to enhance robustness and adaptability. YOLOv9s was evaluated against YOLOv8s and YOLOv5s using standard metrics. It achieved a precision of 86.2%, recall of 88.7%, mAP@0.5 of 90.4%, and mAP@0.5:0.95 of 75.1%, while maintaining a low inference time of 59.3 milliseconds. These results demonstrate YOLOv9s's superior balance between accuracy and computational efficiency. The system supports real-time anomaly detection, making it highly suitable for smart city surveillance and telecom infrastructure monitoring. This work contributes a reliable, scalable solution for enhancing the automation and security of networked environments.

Nomenclature and units

YOLO	You Only Look Once
YOLO-BYTE	You Only Look Once - Binary YouTube Efficient
FODVS	Fiber Optic Detection and Video Surveillance
UAV	Unmanned Aerial Vehicle
PGI	Pole-mounted Ground Infrastructure
SPPF	Spatial Pyramid Pooling - Fast
NMS	Non-Maximum Suppression
FP	False Positive
TP	True Positive
FN	False Negative
IOU	Intersection over Union
AP	Average Precision
PR	Precision-Recall
mAP	mean Average Precision

1. Introduction

The rapid development of telecommunication infrastructure, particularly optical fiber networks, has significantly transformed global communication systems by increasing bandwidth capacity, reducing latency, and enabling high-speed data transmission. As demand for connectivity continues to grow, the deployment of pole-mounted optical fiber infrastructure has become increasingly widespread, especially in regions where underground cabling is economically or logistically unfeasible. These systems are critical for ensuring reliable last-mile connectivity, but they are also inherently vulnerable to environmental and human-induced disturbances (Y. Wang et al., 2019). In particular, their exposure to open environments whether along highways, rural paths, or remote terrains makes them susceptible to a range of threats that can impair system performance and service continuity. Common sources of disruption include animals such as rodents, birds, and larger species that may damage cables through chewing, nesting, or physical interference. Likewise, human activities ranging from theft and vandalism to unauthorized access or illegal construction pose serious risks to infrastructure integrity. Such disturbances not only result in costly maintenance and service interruptions but also raise safety concerns and complicate the logistics of timely intervention (Redmon & Farhadi, 2018). Therefore, there is a growing demand for robust, automated surveillance systems that can identify and classify potential threats in real time and with high accuracy. While traditional surveillance solutions like motion sensors or CCTV systems provide basic monitoring capabilities, they often lack the intelligence to distinguish between types of anomalies or function effectively under varied environmental conditions. These limitations hinder timely threat detection and contribute to false alarms or missed events (Rane, 2023). In response, the application of deep learning-based object detection algorithms has emerged as a promising direction for enhancing infrastructure surveillance. Among these, the You Only Look Once (YOLO) framework stands out for its remarkable balance between detection speed and accuracy, enabling real-time object recognition across complex scenes (Redmon & Farhadi, 2018).

YOLOv9s, a recent lightweight iteration of the YOLO family, is specifically optimized for environments with constrained computational resources while maintaining high precision. It incorporates advanced architectural features such as Programmable Gradient Information (PGI) and GELAN, which enhance feature learning, gradient flow, and multi-scale representation essential for detecting small or partially occluded objects in dynamic outdoor settings (Shi, Li, Liu, Zhou, & Zhou, 2024; C.-Y. Wang, Yeh, & Mark Liao, 2024). These qualities make YOLOv9s particularly suitable for deployment in real-time monitoring systems focused on fiber infrastructure in remote and rural environments. This study explores the use of YOLOv9s to develop an intelligent anomaly detection system that can automatically detect and classify human and animal activity around fiber optic pole installations. The research includes the creation of a custom dataset consisting of annotated real-world images captured under varied weather and lighting conditions along the Ishaka–Kasese road. Through model training, performance benchmarking, and real-time evaluation, this work aims to offer a scalable and efficient solution for improving infrastructure resilience, security, and operational efficiency. Ultimately, the integration of such deep learning systems into fiber optic surveillance holds the potential to transform how telecom networks are monitored, moving toward more automated, intelligent, and proactive infrastructure management strategies. The primary contributions of this paper are as follows:

Developed a custom dataset with a high-resolution camera from the Ishaka-Kasese fiber optic link. The dataset focuses on distinguishing both humans and animals near optical fiber infrastructure installed on a pole in a variety of weather and lighting conditions. Manual annotations were implemented to ensure high labeling accuracy, making the dataset suitable for real-world detection applications. By simulating actual fiber optic surveillance deployment settings, the study records video from independent pole installations, metropolitan areas, and rural areas. This ensures that the model is flexible and applicable to actual challenges in fiber optic infrastructure surveillance, where human and animal intrusions or interferences could compromise service dependability. Comprehensive Assessment Using Conventional Detection Metrics of precision, recall, mAP50, and mAP50-95 are used to evaluate the YOLOv9s model. These metrics offer a comprehensive understanding of the model's ability to manage a range of detection problems, acting as a quantitative benchmark for upcoming research in deep learning-based surveillance-based anomaly detection.

This paper is structured as follows: **Section 2** presents a comprehensive overview of related studies and recent advancements in object detection and surveillance using deep learning techniques. **Section 3** elaborates on the specific methods and main algorithm in this paper. **Sections 4 and 5** discuss the experimental setup, present the results obtained, and provide an in-depth analysis of the model's performance under varying conditions. Finally, **Section 6** offers a conclusive summary of the key findings, and potential directions for future work.

2. Related Study

Recent advancements in deep learning, particularly YOLO-based models, have significantly enhanced real-time detection capabilities in domains such as animal monitoring, infrastructure surveillance, and anomaly detection. Despite their effectiveness, challenges remain regarding model scalability, environmental robustness, and application specificity in critical infrastructure contexts.

2.1 Evolution of YOLO and Surveillance Applications

YOLO's evolution has played a pivotal role in modern object detection. (Redmon & Farhadi, 2018) introduced YOLOv3, which improved detection speed and accuracy, making it suitable for surveillance applications. Subsequent versions, such as YOLOv4 (Ashwin Shenoy & Thillaiarasu, 2023) and YOLOv5 (Pranavan, Aakash, Balakrishnan, Sai, & Cornet, 2023), have been applied to monitor human and animal activity in diverse environments. YOLOv9, the latest iteration, enhances detection through advanced modules like GELAN and PGI, which improve performance on small and occluded objects (Elmir, Touati, & Melizou, 2024). The integration of YOLO with lightweight and hybrid frameworks has led to innovations in surveillance. (Elmir et al., 2024) proposed a hybrid system combining motion detection, frame removal, and YOLOv9, achieving high precision for human and vehicle detection, though it remains susceptible to false alarms in poor weather. Similarly, (Sarker, 2024) proposed an AI-based anomaly detection system with real-time capabilities, supporting intelligent surveillance.

2.2 Infrastructure Monitoring and Anomaly Detection

Multiple studies emphasize deep learning's role in infrastructure monitoring. (Zhang et al., 2025). provided a comprehensive review of surveillance technologies in fiber optic networks, highlighting real-time threat detection mechanisms. (Nguyen et al., 2021) introduce the use of YOLO for rapid identification of risks in critical infrastructure. Sujatha and Janani (2024) extended this by applying deep learning to classify human and animal activities near sensitive installations. (Selim, Hemdan, Shehata, & El-Fishawy, 2021) and (Makled, 2024) explored anomaly detection in large-scale networks using deep models to enhance threat recognition and categorization. (Abdelli et al., 2022) proposed an attention-based BiGRU-autoencoder model for fiber fault detection with high accuracy, though it required significant computational resources and data volume, limiting real-time applicability. Other approaches integrated sensor data with deep learning. (Sha, Feng, Rui, & Zeng, 2021) combined Fiber Optic Distributed Vibration Sensors (FODVS) with YOLOv3 for pipeline event detection, offering a promising fusion of sensing and AI. However, the system remains at a proof-of-concept stage with limited scalability. Similarly, (Andrew, Greatwood, & Burghardt, 2019) used UAVs with YOLOv2 for autonomous species detection, demonstrating potential for non-invasive monitoring, albeit with limited environmental generalization.

2.3 Animal and Behavior Detection in Open Environments

In the domain of animal behavior monitoring, YOLO-based models have shown strong performance but are often confined to agricultural or controlled environments. (Delwar et al., 2025) used YOLOv8 and IoT for real-time animal intrusion detection in farms with 99% accuracy, though performance varied across different animal types. (Chan et al., 2024) introduced YOLO-Behaviour, enabling multi-species behavior recognition with minimal training data, but the system needed customization for complex behavioral patterns. (Zheng, Li, & Qin, 2023) further advanced livestock monitoring using a YOLO-BYTE hybrid, improving tracking but requiring careful camera placement. (Lu, Li, & Lu, 2024) developed an objectness-aware network for wildlife detection in dense habitats, which improved localization but struggled with occlusion and low-light scenarios. (Payghode, Goyal, Bhan, Iyer, & Dubey, 2023) focused on activity recognition in video surveillance using neural networks but noted high processing costs for real-time applications.

2.4 Research Gap and Justification

While numerous studies highlight the efficacy of YOLO-based models in surveillance and animal monitoring, they often lack contextual tailoring to pole-mounted fiber optic infrastructure. For instance, although (Elmir et al., 2024) and (Payghode et al., 2023) demonstrated strong general surveillance performance, their solutions were not adapted for telecom-specific environments. Similarly, fiber-specific sensing approaches like that of (Abdelli et al., 2022) omit real-time classification of human and animal threats. Additionally, animal detection studies (Chan et al., 2024; Delwar et al., 2025) are predominantly based in farm settings and do not address the unique risks posed in rural or semi-urban telecom deployments, such as tampering or unauthorized access. This creates a gap in the development of domain-specific surveillance systems. This study addresses that gap by designing a YOLOv9s-based system specifically tailored to detect and classify human and animal activity near pole-mounted optical fiber infrastructure. By focusing on real-world, open-environment scenarios relevant to telecom networks, this research contributes a specialized framework that improves surveillance granularity and enhances infrastructure resilience and predictive maintenance, especially in remote and resource-limited regions.

3. Methodology

3.1. Overview of YOLOv9s of Model

YOLOv9s is a compact, efficient object detection model that was released as part of the YOLOv9 family. It is primarily designed for real-time deployment on edge devices with low processing resources. It strikes an excellent mix between accuracy, speed, and model size, making it ideal for applications such as surveillance, smart monitoring, and infrastructure security

(Yaseen, 2024). Compared to previous lightweight variations like YOLOv5n and YOLOv8n, YOLOv9s takes advantage of modern architectural advances while keeping a low computational footprint (Zhang et al., 2025). This makes it exceptional for circumstances requiring rapid detection of small or moving objects, such as human and animal movement, in order for the system to respond. The YOLOv9s variation combines recent training and inference optimizations while maintaining YOLO's signature single-stage detection structure. It offers quantization, pruning, and numerous deployment formats, making it simple to integrate into a wide range of systems, including embedded surveillance units and remote sensing modules (Jin Li, Gong, Hou, & Wang, 2024). With its increased speed and detection precision, YOLOv9s stands out as an ideal solution for security-critical infrastructure monitoring, particularly in outdoor or rural locations where computing efficiency and model portability are key.

3.1.1. Analysis of the Original YOLOv9s Network Structure

The YOLOv9s model is a lightweight yet capable form of the YOLOv9 object-detection series, specifically designed to reconcile high detection precision with computing efficiency (C.-Y. Wang et al., 2024). YOLOv9s builds on the legacy of its predecessors in the YOLO series, including a number of architectural changes and training methodologies that considerably improve its performance in real-time image identification and object detection tasks (Yaseen, 2024). YOLOv9 is built around the Programmable Gradient Information (PGI) framework, a unique technique that ensures precise gradient flow during backpropagation. This approach increases the quality of gradient information used in the computation of the objective function, allowing for more effective network weight updates. In reality, PGI leads to faster convergence during training and higher overall detection precision, especially in difficult circumstances with occlusions, overlapping objects, or variable object scales (An, Zhang, Sun, & Wang, 2024). Figure 1 present the structure of the YOLOv9s object detection network, including the data flow from input to output. Convolutional (Conv) layers are used in the early stages of processing to extract important low-level information from input images. These layers record patterns including edges, textures, and basic shapes, which serve as the foundation for more in-depth semantic analysis. The network then moves on to a more advanced feature processing stage with the addition of the RepNCSPELAN4 module, which serves as the primary feature extraction and fusion unit (Shi et al., 2024; Vo, Mui, Thien, & Tien, 2024). This module is intended to aggregate characteristics from various depths of the network, effectively combining spatial and semantic information to improve the robustness of object representation (Jun Li, Feng, Shao, & Liu, 2024). To enable multi-scale object detection, the Concatenation (Concat) technique is used to combine feature maps from several levels of the network.

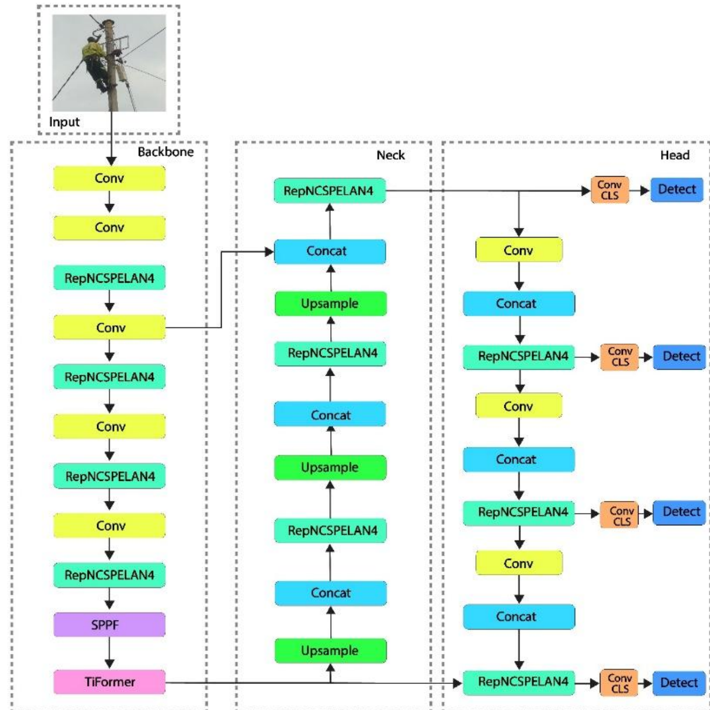

Figure 1: The Structure of Yolov9s Model

This technique allows for the integration of both fine-grained and abstract information, which is especially important when recognizing objects of variable sizes, such as small animals or partially visible intruders in surveillance film. Following that, the Spatial Pyramid Pooling - Fast (SPPF) module is added to improve the model's ability to deal with scale fluctuations. The SPPF module concatenates information from several spatial scales, allowing the model to understand contextual cues as well as local features. This multi-scale awareness improves the accuracy of recognizing both close and distant targets in complex real-world contexts. In the final stage of detection, the identified objects are classified within their bounding boxes using a Convolutional Classification Layer (Conv CLS). Each bounding box is allocated a class label based on the learnt attributes. The Detect layer generates the final detection output, converting the classifications and bounding box coordinates into actionable findings. It also uses Non-Maximum Suppression (NMS) to reduce duplicate detections by keeping only the most confident bounding box for overlapping predictions. Throughout the design, directional arrows depict the flow of data, beginning with the input image and progressing sequentially via convolutional feature extraction, multi-scale fusion, and classification, culminating in the final, refined object detection output.

4. Experimental Environment and Dataset Setup

4.1. Dataset Description

The dataset contains a broad collection of images that shows human and animal activity around pole-mounted optical fiber infrastructure, especially in rural and semi-urban areas. It contains around 1500 annotated instances of different objects, such as humans (e.g., technicians, pedestrians) and animals (e.g., monkeys, chimpanzee, birds). The annotations contain bounding boundaries and activity labels, allowing for fine-grained object classification and activity detection. To improve model robustness, the dataset is divided into three sets: training (70%), validation (20%), and testing (10%), and enriched with techniques such as flipping, brightness variation, and cropping to simulate real-world environmental variability. The use of coarse-grained labels "human," "animal," and "pole" was a strategic decision grounded in the need to balance system performance, annotation efficiency, and contextual relevance for infrastructure surveillance. These categories directly support the study's objectives: detecting human intrusion, animal interference, and monitoring structural elements critical to pole-mounted optical fiber systems. Finer-grained labels could increase annotation effort, introduce classification ambiguity, and raise computational demands, thereby compromising real-time detection performance. Given the deployment context often remote or resource-constrained areas maintaining low latency and high accuracy was prioritized. Thus, the adopted labeling strategy ensures operational effectiveness while avoiding unnecessary complexity, aligning with the system's intended use for anomaly detection and preventive maintenance in challenging environments. The datasets and code used in this study are available from the corresponding author upon reasonable request.

4.2. Dataset Production

To produce a diversified, high-quality, and representative dataset suited for robust model training and evaluation, data collection was thoroughly planned and carried out in a variety of environmental scenarios and geographical regions. The random photos were taken along the Ishaka-Kasese route, with the goal of identifying and capturing both human and animal movement near fiber optic poles. The data collection procedure comprised recordings at different times of day morning, afternoon, and evening to capture variations in natural illumination and movement patterns. Furthermore, footage was captured in a variety of weather circumstances, including sunny, cloudy, and rainy settings, to approximate real-world unpredictability. Data was collected in a variety of applications including urban areas, rural communities, and remote pole sites, to guarantee broad contextual representation. infrastructure. This extensive method greatly increased the dataset's diversity and richness, which improved the YOLOv9s model's generalizability across various security and maintenance surveillance scenarios involving optical fiber infrastructure.

4.3. Image Acquisition and Labeling Using YOLOv9s

The obtained images were manually annotated using LabelImg, an open-source graphical annotation tool commonly utilized in object detection applications. This labeling method was critical in training the YOLOv9s model to reliably recognize and classify key objects such as poles, human, animals, and specific activities occurring around pole-mounted fiber infrastructure. Each object of interest was wrapped in a precise bounding box to offer the YOLOv9s-model with the detailed positional and spatial information it requires for accurate localization and classification. Bounding boxes were carefully designed to snugly contain the target objects while decreasing background noise, hence boosting the model's capacity to acquire reliable feature representations. The annotations were saved in YOLO format, which is compatible with the YOLOv9 training pipelines, and were designed to facilitate multi-class detection and fine-grained activity recognition. The annotations were saved in YOLO format, which is compatible with the YOLOv9 training pipelines, and were designed to facilitate multi-class detection and fine-grained activity recognition. Table 1 present the class frequency and distribution for the annotated images

Table 1: The class frequency and distribution for the annotated images

Class Name	Number of Instances	Percentage of Total (%)
Human	750	50.0%
Animal	500	33.3%
Pole	250	16.7%
Total	**1500**	**100%**

4.4. Class Label Assignment and Dataset Consistency

The class labels were assigned using a preset schema that divided objects and actions into various groups related to the study's monitoring aims. The basic classes were human, animal, and pole. Each label was awarded based on visual qualities and contextual information in the photograph, ensuring that even minor differences in posture or behavior were accurately caught. To ensure clarity and minimize ambiguity, a tight class name strategy was followed throughout all annotations. Several quality tests were carried out on the labeled images to ensure consistency and dependability in the dataset. Cross-validation was utilized by annotators to detect mislabeled items, irregular bounding box sizes, and class borders that overlapped. Any inconsistencies or errors were addressed in cooperation with a lead annotator to ensure that annotation standards were followed. Furthermore, annotation rules were defined and given to all labelers in order to reduce subjective interpretation, particularly for activity-based labels with movement cues that could be misunderstood. These efforts ensured that every occurrence in the dataset was labeled with high accuracy and consistency, which is necessary for optimizing the YOLOv9s model's learning process. The dataset was also assessed for class balance and distribution to verify that no one group was either overrepresented or under-represented, as this could create bias during training. In circumstances where certain activity classes like human, animal and pole were under-sampled, targeted data augmentation was done to artificially expand these subsets. This balance ensured that the YOLOv9s model could generalize well across all classes while also performing robustly in real-time surveillance applications. Consistent labeling, unambiguous class definitions, and balanced representation resulted in a high-quality dataset appropriate for developing an efficient and accurate YOLOv9s-based detection system.

4.5. Training Configuration and Augmentation Techniques Using YOLOv9s

The YOLOv9s model was trained and evaluated on a workstation equipped with an Intel Core i7-6700HQ CPU running at 2.60 GHz, 24 GB of RAM, and a 500 GB SSD, ensuring a stable and efficient computational setup. A NVIDIA Quadro 1000M GPU was used to accelerate model training via its CUDA cores, while Python 3.12 served as the primary programming interface. The training environment was managed through Anaconda and implemented using the PyTorch deep learning framework, which offered full support for the YOLOv9 architecture. Input images were resized to 640 × 640 pixels, aligning with YOLOv9s' optimal detection resolution. The model was trained for 100 epochs using a batch size of 16 and an initial learning rate of 0.001, with the Adam optimizer employed to update model parameters. Model checkpoints were saved periodically to the SSD to ensure rapid loading and recovery. A comprehensive data augmentation strategy was applied to enhance the model's generalizability. This included random rotation, scaling, and cropping, which simulated variations in object angle, distance, and partial occlusions. These augmentations were particularly important for modeling real-world conditions encountered in surveillance of pole-mounted optical fiber infrastructure, improving the model's ability to detect and classify human and animal activities under diverse scenarios.

4.6. Evaluation Metrics

In order to evaluate how well object detection models identify and categorize human and animal activity around pole-mounted optical fiber infrastructure, this study uses four commonly used evaluation metrics: mean Average Precision at IoU threshold 0.5 (mAP@0.5), mean Average Precision across IoU thresholds from 0.5 to 0.95 (mAP@0.5:0.95), Precision (P), and Recall (R). Precision is defined as the fraction of accurately predicted positive samples (true positives) compared to all positive samples predicted. It shows the model's capacity to minimize false alarms, which is important in surveillance applications because incorrect detections might result in irrelevant alerts or interventions.

Recall assesses the model's ability to identify all true positive cases. It is defined as the ratio of true positives to total positives (true positives plus false negatives). High recall suggests that the model is effective at capturing all important events, such as human invasions or animal activity that may interfere with fiber infrastructure. The mAP@0.5 acts as a baseline for the model's detection accuracy. It is determined by taking the average of the average accuracy (AP) scores for each class and applying a fixed Intersection over Union (IoU) threshold of 0.5. This statistic measures how successfully the model localizes and identifies items with a reasonable overlap margin. For a more rigorous and complete assessment, the study employs mAP@0.5:0.95, which averages AP scores over 91 IoU criteria ranging from 0.5 to 0.95 in 0.05 increments. This statistic provides a more detailed understanding of the model's robustness under changing localization tolerances, making it especially useful in real-world scenarios where objects may be partially obscured, in motion, or vary in size and visibility.

The mathematical definitions of these metrics are shown from equation (1) to (4):

$$P = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

$$R = \frac{\text{TP}}{\text{TP+FN}} \tag{2}$$

$$mAP_{50} = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{3}$$

$$mAP_{50-95} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{91}\sum_{j=i}^{91} AP_{ij}\right) \tag{4}$$

Where: TP represent the number of true positive detections, FP is the number of false positives, FN is the number of false negatives, n is the number of classes, AP_i is the average precision for class iii at an IoU threshold of 0.5, and AP_{ij} represents the average precision for class i at IoU thresholds from 0.5 to 0.95, incremented by 0.005.

5. Results

5.1. Analysis of YOLOv9s Precision-Recall and Confidence-Based Performance Curves

Figure 2 shows four essential performance graphs for the YOLOv9s model: the Precision-Recall (PR) curve, Precision-Confidence curve, F1-Confidence curve, and the Recall-

Confidence curve. The PR curve demonstrates the trade-off between precision and recall at various confidence levels. A robust PR curve for YOLOv9s is seen, with high accuracy values maintained even when recall increases, indicating that the model can consistently detect true positives while limiting false positives. This smooth, high-arching PR curve demonstrates YOLOv9s' robustness in detecting complicated human and animal behaviors near pole-mounted optical fiber infrastructure. The area under the PR curve (AUC-PR) is large, indicating that the model can generalize well to new data and is not unduly sensitive to changes in detection thresholds. The Precision-Confidence curve reinforces the model's reliability. As the confidence score rises, the model retains high precision, peaking at confidence levels greater than 0.7. This indicates that when YOLOv9s produces predictions with high confidence, they are highly likely to be correct an important feature for real-time surveillance applications where false alarms must be reduced. The F1-Confidence curve shows that the F1-score a harmonic mean of precision and recall reaches its peak between 0.5 and 0.7 confidence, indicating that this range is the best threshold for balanced performance. This peak implies that YOLOv9s is both precise and efficient in minimizing missed detections and false positives, resulting in a well-calibrated model appropriate for automated anomaly detection.

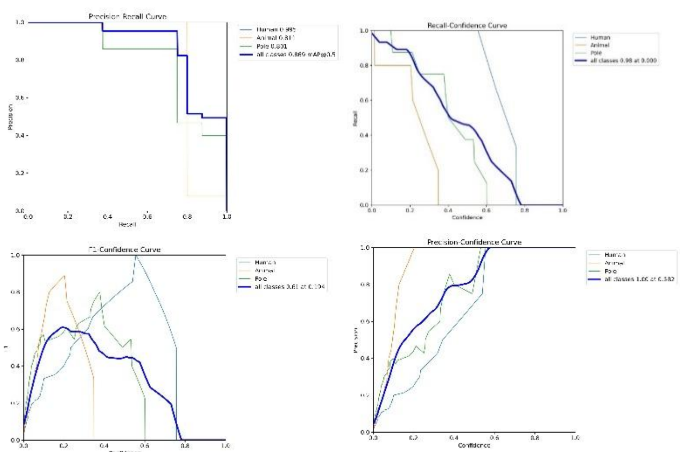

Figure 2: the precision-recall curve, the precision-confidence curve, the F1-confidence curve, and the recall-confidence curve.

Finally, the Recall-Confidence curve shows how the recall score changes as the confidence threshold varies. YOLOv9s retains strong recall values across a wide confidence range, with only a tiny drop at a confidence level of 0.8. This suggests that the model will continue to identify the majority of relevant items even when the decision threshold is raised. High recall at lower and mid-range confidence levels confirms that important events, such as human presence or animal activity, are not left behind. Together, these performance curves confirm YOLOv9s as an accurate, dependable, and well-optimized model capable of offering excellent performance in security and maintenance surveillance systems, particularly for intelligent monitoring of pole-mounted fiber-optic installations.

5.2. Comparative Analysis of Training and Validation Metrics for YOLOv9s, YOLOv8s, and YOLOv5s

Figure 3 provides a comparative evaluation of the training and validation performance of three object detection modelsYOLOv9s, YOLOv8s, and YOLOv5s trained over 100 epochs on a custom dataset. The subplots illustrate trends in critical performance metrics, including training and validation losses, precision, recall, and mean Average Precision (mAP). These indicators are essential for understanding each model's convergence behavior, generalization ability, and detection accuracy. YOLOv9s, in particular, exhibits a smoother and more consistent decline in both loss curves, suggesting effective learning without premature overfitting. Throughout training, YOLOv9s consistently outperforms the other models in precision and recall. Its recall approaches 0.90 by the final epochs, indicating robust capability in identifying true positives. Simultaneously, the model maintains high precision, reflecting a low incidence of false positives. In contrast, YOLOv8s and YOLOv5s demonstrate more fluctuating performance, with lower precision and recall scores that reflect less reliable object differentiation. These results affirm YOLOv9s's enhanced detection consistency across various classes, including poles, humans, and animals.

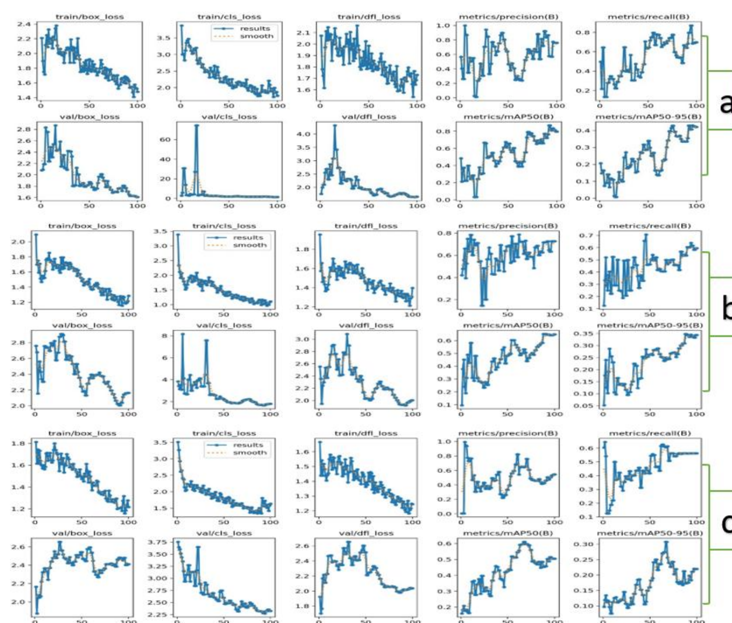

Figure 3: (a) The training results and validation performance metrics of YOLOv9s at 100 epochs. (b) The training results and validation performance metrics of YOLOv8s at 100 epochs. (c) The training results and validation performance metrics of YOLOv5s at 100 epochs.

In terms of mAP@0.5, YOLOv9s achieves scores exceeding 0.80, highlighting its superior ability to localize and classify objects under varying spatial and contextual conditions. This advantage stems from improvements in network architecture, label assignment strategy, and feature extraction pipelines, which collectively enhance learning efficiency. While YOLOv8s and YOLOv5s show incremental gains, their performance plateaus earlier, constrained by older architectural frameworks. Consequently, YOLOv9s emerges as the most effective model for integration into the proposed intelligent surveillance framework, particularly for applications requiring real-time, high-accuracy monitoring of pole-mounted fiber-optic infrastructure.

5.3. Detection Results Analysis

Following training on a customized dataset designed for monitoring pole-mounted optical fiber infrastructure, Figure 4 shows a selection of detection results from the YOLOv9s model. The model is highly accurate in distinguishing various item classes, including individuals, animals, and poles, across a wide range of environmental and lighting circumstances. Bounding boxes are drawn precisely around the discovered items, with class labels and confidence ratings clearly shown. This demonstrates the model's capacity to distinguish between different activity categories with little misunderstanding, which is essential for real-time surveillance applications. The visual results show that YOLOv9s can detect partial occlusions, variable object scales, and non-standard postures, such as a person mounting a pole, which is particularly useful for intrusion detection and infrastructure safety monitoring. Furthermore, the model performs well in reducing false positives while properly identifying actual cases in both high-contrast and chaotic backdrops. Detection is consistent across angles and distances, demonstrating the usefulness of data augmentation approaches and the model architecture's durability. The confidence scores reported in the bounding boxes are consistently within a reliable range (often above 0.80), suggesting the model's high confidence in its predictions. These qualitative findings are consistent with the evaluation metrics (mAP, Precision, and Recall) stated earlier, indicating that YOLOv9s is well-suited for use in intelligent monitoring systems intended at improving the security and operational integrity of outdoor fiber-optic networks.

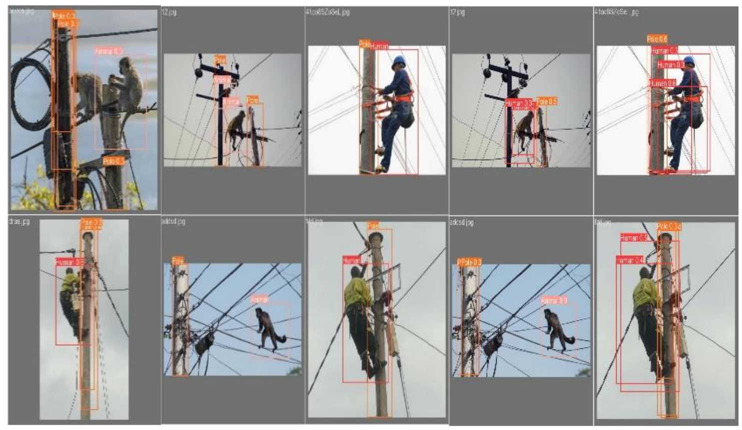

Figure 4: Detection results of the trained dataset

5.4. Performance Comparison Analysis of YOLOv9s, YOLOv8s, and YOLOv5s Models

Table 2 provides a complete comparison of the performance of three object detection models YOLOv5s, YOLOv8s, and the more advanced YOLOv9s on the same unique dataset used for surveillance of pole-mounted optical fiber infrastructure. YOLOv9s surpasses its predecessors on every performance metric. In terms of precision, YOLOv9s achieves 86.2%, suggesting that it has the lowest rate of false positives, or inaccurate predictions regarding the presence of objects. YOLOv8s and YOLOv5s follow with 81.4% and 78.5% accuracy, respectively. This increased trend in precision demonstrates the gains in detection accuracy of newer YOLO models, which are mostly attributable to advancements in model architecture and feature extraction capabilities.

The recall score, which measures the model's ability to detect all relevant objects, is again greatest in YOLOv9s, at 88.7%, outperforming YOLOv8s (84.3%) and YOLOv5s (81.6%). This high recall indicates that YOLOv9s can detect almost all occurrences of poles, humans, and animals activities without omission. Furthermore, the mean Average Precision (mAP@0.5), a crucial criterion for detection performance, reaches 90.4% for YOLOv9s, beating YOLOv8s (87.2%) and YOLOv5. Furthermore, in the more stringent mAP@0.5:0.95 metric, YOLOv9s leads with 75.1%, followed by YOLOv8s at 71.5% and YOLOv5s at a significantly lower 57.2%. This demonstrates YOLOv9s' greater capacity to reliably identify items across varied degrees of overlap (IoU thresholds), making it ideal for complicated, real-world situations.

Table 2: The comparison with Other YOLO Models

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Inference Time (ms)
YOLOv5s	78.5	81.6	86.4	57.2	80.5
YOLOv8s	81.4	84.3	87.2	71.5	73.2
YOLOv9s	86.2	88.7	90.4	75.1	59.3

Another considerable improvement is observed in inference time, with YOLOv9s having the fastest processing speed at 59.3 ms per image, compared to YOLOv8s at 73.2 ms and YOLOv5s at 80.5 ms. This latency reduction is critical for real-time deployment in security and surveillance systems, where rapid detection and response are required. YOLOv9s' efficiency is due to its optimized backbone, transformer integration, and improved feature fusion modules, which reduce computational overhead while maintaining accuracy. Overall, the comparison research shows that YOLOv9s not only has the best detection performance but also provides faster and more resource-efficient inference, making it the best choice for intelligent monitoring systems in dynamic outdoor environments.

5.5. Effect of Training Epochs on YOLOv9s Model Accuracy

Table 3 shows how the performance of the YOLOv9s model improves as the number of training epochs grows, shedding light on the model's learning behavior over time. After 20 epochs, the model has a precision of 80.7%, a recall of 73.5%, and a mAP@0.5 of 77.8%. While these initial numbers indicate that the model is learning to recognize object features and associations in the dataset, the comparatively low recall and mAP@0.5:0.95 (58.4%) indicate that more training is required to capture more complex patterns and reduce misdetections. The performance at this stage reflects the model's early phases of learning, when it still misses numerous true items and has not yet generalized successfully. As the training advances to 50 and then 100 epochs, significant improvements are seen across all measures. At 50 epochs, the precision rises to 82.8% and the recall to 84.1%, indicating that the model is successfully learning to recognize more relevant objects with fewer false positives. The mAP@0.5 increases significantly to 87.6%, whereas the more severe mAP@0.5:0.95 improves to 70.5%, indicating improved localization accuracy across several IoU thresholds. By 100 epochs, YOLOv9s has achieved peak performance with 86.2% precision, 88.7% recall, 90.4% mAP@0.5, and 75.1% mAP@0.5:0.95.

Table 3: Effect of Training Epochs on Model Accuracy

Epochs	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
20	80.7	73.5	77.8	58.4
50	82.8	84.1	87.6	70.5
100	86.2	88.7	90.4	75.1

These findings demonstrate that longer training enables the model to increase its feature extraction, generalization, and detection performance. The learning curve shows that YOLOv9s improves significantly from lengthier training, making it ideal for high-accuracy detection tasks in real-world settings.

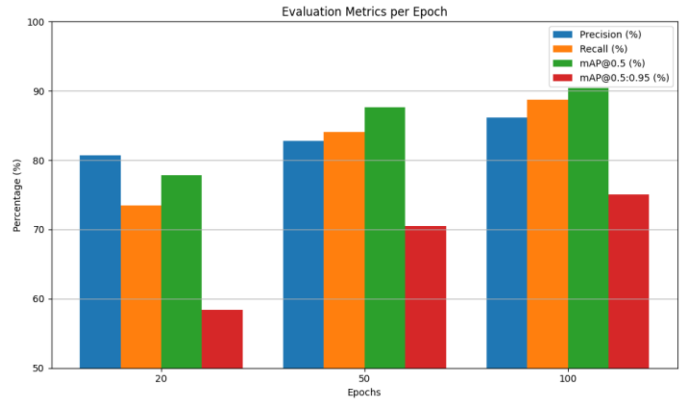

Figure 5: The Evaluation Metrics Across Training Epochs

Figure 5 presents the progression of four key evaluation metrics Precision, Recall, mAP@0.5, and mAP@0.5:0.95 across training epochs 20, 50, and 100. As training advances, all metrics show a clear upward trend, reflecting improved model performance. Precision rises from 80.7% to 86.2%, while Recall increases more significantly from 73.5% to 88.7%, suggesting better detection capabilities. Similarly, mAP@0.5 grows from 77.8% to 90.4%, and mAP@0.5:0.95 improves from 58.4% to 75.1%, indicating enhanced accuracy over both easy and challenging detection thresholds. Overall, the chart demonstrates that extended training substantially boosts the model's effectiveness.

6. Conclusion

This study effectively confirmed the successful results of the YOLOv9s model in detecting anomalies around fiber optic infrastructure, notably in pole-mounted scenarios using a complete and methodical deep learning approach. The primary goal of improving intelligent video surveillance for infrastructure safety was realized by exploiting YOLOv9s' significant architectural upgrades and detection capabilities. The model was trained and tested on a well-curated, annotated dataset, with an emphasis on dataset consistency, class label assignment, and robust augmentation strategies to improve generalization across a range of real-world scenarios. A notable accomplishment of this work is the evident performance superiority of YOLOv9s over its predecessors, YOLOv8s and YOLOv5. YOLOv9s achieved an impressive precision of 86.2%, a recall of 88.7%, and a high mAP@0.5 of 90.4%, with mAP@0.5:0.95 reaching 75.1%. These measurements illustrate the model's ability to reliably detect and categorize a wide range of object types, including poles, human, and animals' motions, as well as its constant performance over different localization criteria. Furthermore, the model maintained a competitive inference time of 59.3 ms, demonstrating its appropriateness for real-time deployment in intelligent surveillance systems. Another notable addition of this work is the systematic training configuration and implementation of targeted augmentation procedures, which considerably improved model robustness. The precision-recall and confidence curves supported YOLOv9s' ability to balance detection accuracy and confidence scores. Finally, the study confirms YOLOv9s as a highly effective approach for automated anomaly detection in fiber optic monitoring systems. Its excellent accuracy, efficiency, and real-time performance make it a feasible model for implementation in smart infrastructure protection applications, opening the door to more adaptable and intelligent security solutions in the telecom sector and beyond.

Future research will aim to systematically expand the dataset by incorporating diverse environmental conditions (e.g., rain, fog, and varying daylight), multiple camera perspectives, and a broader taxonomy of human and animal activities. This is hypothesized to enhance the model's robustness and generalization across heterogeneous real-world scenarios. A second research objective involves integrating night vision and thermal infrared imaging modalities to empirically evaluate improvements in detection accuracy under low-light or visually obstructed conditions. Additionally, lightweight optimization techniques such as model quantization, pruning, and knowledge distillation will be explored to facilitate efficient real-time inference on edge computing devices embedded within fiber-optic surveillance networks. Lastly, the inclusion of temporal sequence modeling (e.g., using ConvLSTM or transformer-based video encoders) will be investigated to assess its potential in improving activity recognition and anomaly detection over time, thus supporting the development of intelligent, context-aware monitoring systems.

References

Abdelli, K., Cho, J. Y., Azendorf, F., Griesser, H., Tropschug, C., & Pachnicke, S. (2022). Machine-learning-based anomaly detection in optical fiber monitoring. *Journal of optical communications and networking, 14*(5), 365-375.

An, R., Zhang, X., Sun, M., & Wang, G. (2024). GC-YOLOv9: Innovative smart city traffic monitoring solution. *Alexandria Engineering Journal, 106*, 277-287.

Andrew, W., Greatwood, C., & Burghardt, T. (2019). *Aerial animal biometrics: Individual friesian cattle recovery and visual identification via an autonomous uav with onboard deep inference.* Paper presented at the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Ashwin Shenoy, M., & Thillaiarasu, N. (2023). Enhancing temple surveillance through human activity recognition: A novel dataset and YOLOv4-ConvLSTM approach. *Journal of Intelligent & Fuzzy Systems, 45*(6), 11217-11232.

Chan, A. H. H., Putra, P., Schupp, H., Köchling, J., Straßheim, J., Renner, B., . . . Burke, T. (2024). YOLO-Behaviour: A simple, flexible framework to automatically quantify animal behaviours from videos. *Methods in Ecology and Evolution*.

Delwar, T. S., Mukhopadhyay, S., Kumar, A., Singh, M., Lee, Y.-w., Ryu, J.-Y., & Hosen, A. S. (2025). Real-Time Farm Surveillance Using IoT and YOLOv8 for Animal Intrusion Detection. *Future Internet, 17*(2), 70.

Elmir, Y., Touati, H., & Melizou, O. (2024). Intelligent Video Recording Optimization using Activity Detection for Surveillance Systems. *arXiv preprint arXiv:2411.02632*.

Li, J., Feng, Y., Shao, Y., & Liu, F. (2024). IDP-YOLOV9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective. *Applied Sciences, 14*(12), 5277.

Li, J., Gong, W., Hou, L., & Wang, J. (2024). *Lightweight object detection algorithm based on improved YOLOv8n.* Paper presented at the Proceedings of the 2024 6th International Conference on Video, Signal and Image Processing.

Lu, X., Li, W., & Lu, X. (2024). An objectness-aware network for wildlife detection. *Multimedia Tools and Applications, 83*(3), 7119-7133.

Makled, E. A. M. M. (2024). *Advanced optimization and machine learning techniques for efficient wireless communication networks.* Memorial University of Newfoundland,

Nguyen, H. H., Ta, T. N., Nguyen, N. C., Bui, V. T., Pham, H. M., & Nguyen, D. M. (2021). *Yolo based real-time human detection for smart video surveillance at the edge.* Paper presented at the 2020 IEEE eighth international conference on communications and electronics (ICCE).

Payghode, V., Goyal, A., Bhan, A., Iyer, S. S., & Dubey, A. K. (2023). Object detection and activity recognition in video surveillance using neural networks. *International journal of web information systems, 19*(3/4), 123-138.

Pranavan, V., Aakash, S., Balakrishnan, S., Sai, V., & Cornet, S. (2023). *Integrating technology, local knowledge, and participatory approach: real-time detection system for mitigating human-animal conflicts in rural India.* Paper presented at the 2023 IEEE Global Humanitarian Technology Conference (GHTC).

Rane, N. (2023). YOLO and Faster R-CNN object detection for smart Industry 4.0 and Industry 5.0: applications, challenges, and opportunities. *Available at SSRN 4624206*.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Sarker, I. H. (2024). AI for critical infrastructure protection and resilience. *AI-driven cybersecurity and threat intelligence: Cyber automation, intelligent decision-making and explainability*, 153-172.

Selim, G. E. I., Hemdan, E. E.-D., Shehata, A. M., & El-Fishawy, N. A. (2021). Anomaly events classification and detection system in critical industrial internet of things infrastructure using machine learning algorithms. *Multimedia Tools and Applications, 80*(8), 12619-12640.

Sha, Z., Feng, H., Rui, X., & Zeng, Z. (2021). PIG tracking utilizing fiber optic distributed vibration sensor and YOLO. *Journal of Lightwave Technology, 39*(13), 4535-4541.

Shi, Y., Li, S., Liu, Z., Zhou, Z., & Zhou, X. (2024). MTP-YOLO: You only look once based maritime tiny person detector for emergency rescue. *Journal of Marine Science and Engineering, 12*(4), 669.

Vo, H.-T., Mui, K. C., Thien, N. N., & Tien, P. P. (2024). Automating tomato ripeness classification and counting with YOLOv9. *Int. J. Adv. Comput. Sci. Appl, 15*(4).

Wang, C.-Y., Yeh, I.-H., & Mark Liao, H.-Y. (2024). *Yolov9: Learning what you want to learn using programmable gradient information.* Paper presented at the European conference on computer vision.

Wang, Y., Yuan, H., Liu, X., Bai, Q., Zhang, H., Gao, Y., & Jin, B. (2019). A comprehensive study of optical fiber acoustic sensing. *IEEE access, 7*, 85821-85837.

Yaseen, M. (2024). What is yolov9: An in-depth exploration of the internal features of the next-generation object detector. *arXiv preprint arXiv:2409.07813*.

Zhang, X.-Y., Xuan, Y., Mu, C., Ding, Z., Wang, H., & Guo, P. (2025). Robust Lightweight UAV Inspection System for Consumer Electronics Applications in Smart Grids. *IEEE Transactions on Consumer Electronics*.

Zheng, Z., Li, J., & Qin, L. (2023). YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Computers and Electronics in Agriculture, 209*, 107857.