Hardware-Accelerated Detection of Unauthorized Mining Activities Using YOLOv11 and FPGA

Refka Ghodhbani[1], Taoufik Saidani[2]*, Amani Kachoukh[3],
Mahmoud Salaheldin Elsayed[4], Yahia Said[5], Rabie Ahmed[6]
Center for Scientific Research and Entrepreneurship, Northern Border University, 73213, Arar, Saudi Arabia[1, 2]
Department of Information Systems-Faculty of Computing and Information Technology,
Northern Border University, Saudi Arabia[3]
Department of Computer Sciences-Faculty of Computing and Information Technology,
Northern Border University, Saudi Arabia[4]
DCenter for Scientific Research and Entrepreneurship, Northern Border University, 73213, Arar, Saudi Arabia[5]
Department of Computer Science-Faculty of Computing and Information Technology,
Northern Border University, Rafha, Saudi Arabia[6]
Mathematics and Computer Science Department-Faculty of Science, Beni-Suef University, Beni-Suef, Egypt[6]

Abstract—**Illegal mining activities present significant environmental, economic, and safety challenges, particularly in remote and under-monitored regions. Traditional surveillance methods are often inefficient, labor-intensive, and unable to provide real-time insights. To address this issue, this study proposes a computer vision-based solution leveraging the state-of-the-art YOLOv11 Nano and Small models, fine-tuned for the detection of illegal mining activities. A specific dataset comprising aerial and ground-level images of mining sites was curated and annotated to train the models for identifying unauthorized excavation, equipment usage, and human presence in restricted zones. The proposed system integrates the hardware-software design of YOLOv11 on the PynqZ1 FPGA, offering a high-performance, low-latency, and energy-efficient solution suitable for real-time monitoring in resource-constrained environments. This hardware-accelerated approach combines FPGA's parallel processing capabilities with the lightweight deep learning models, enabling efficient deployment for automated illegal mining detection. By providing a scalable, real-time monitoring tool, this work contributes to the development of automated enforcement tools for the mining industry, ensuring better control and surveillance of mining activities. To validate the efficiency of deep learning deployment on edge devices, YOLOv11n was implemented on an FPGA, utilizing 70% of available LUTs, 50% of FFs, and 80% of DSPs, with 8.3 Mbits of on-chip memory. The design achieved 100.33 GOP/s throughput, 18 FPS at 55 ms latency, consuming 4.8 W, and delivering an energy efficiency of 20.90 GOP/s/W.**

Keywords—YOLOv11; object detection; mining industry

I. Introduction

Illegal mining represents a pressing and multifaceted global issue that continues to challenge environmental governance, economic stability, and social equity across both developed and developing regions. The unsanctioned and unregulated extraction of mineral resources leads to significant financial losses for national governments by circumventing taxation systems, depleting natural capital, and enabling the growth of informal markets [1], [2], [3]. The widespread prevalence of illegal mining has been particularly damaging in regions rich in natural resources, such as parts of Africa, South America, and Southeast Asia, where limited institutional oversight and socio-economic vulnerabilities contribute to the proliferation of these activities.

From an environmental perspective, illegal mining contributes to extensive and often irreversible ecological degradation. It leads to deforestation, soil destabilization, and contamination of surface and groundwater resources through the release of heavy metals and toxic chemicals like mercury, arsenic, and cyanide [2], [3]. These pollutants have long-lasting consequences on local biodiversity and human health, often affecting downstream communities that rely on natural water sources. Furthermore, land surface changes caused by mining disrupt natural drainage patterns and increase the risk of landslides, sedimentation, and flooding, compounding the environmental impact in fragile ecosystems.

Socially, illegal mining exacerbates inequality, fuels conflict, and often involves exploitative labor practices. Workers in illegal mines typically operate without protective equipment or health and safety protocols, exposing them to life-threatening conditions such as tunnel collapses, toxic exposure, and physical abuse [4]. Child labor is also a recurring issue in illegal mining operations, raising serious human rights concerns. Moreover, these activities are frequently linked to criminal networks, including trafficking, corruption, and violent conflict over territorial control. The lack of regulation and oversight creates a fertile ground for systemic abuse and contributes to broader instability within affected communities.

Despite the severity of these impacts, monitoring and controlling illegal mining remain formidable challenges for governments and international organizations. Traditional methods, such as field inspections, aerial surveys, and manual satellite image interpretation, are limited in scope, costly to implement, and incapable of providing continuous, real-time monitoring [5], [6]. These methods often suffer from temporal lags and spatial blind spots, especially in remote, forested, or mountainous regions where illegal mining thrives under the radar. Furthermore, these conventional systems often rely on human expertise for image analysis, making them susceptible to errors, biases, and inconsistencies in detection.

*Corresponding authors.

In recent years, technological advancements in remote sensing, machine learning, and computer vision have opened new possibilities for addressing the limitations of traditional monitoring systems. The integration of satellite imagery with automated analysis tools, particularly deep learning models, has demonstrated strong potential for detecting and localizing mining activity in diverse environments [7], [8]. High-resolution Earth observation platforms, such as Sentinel and Landsat, have made large-scale environmental monitoring more accessible, while the growing availability of labeled datasets has enabled the training of powerful object detection models capable of identifying complex patterns and features associated with illegal mining operations.

Among the various object detection frameworks, the YOLO (You Only Look Once) architecture has gained prominence due to its remarkable trade-off between speed and accuracy. Recent iterations of YOLO, such as YOLOv5, YOLOv8, and the newer YOLOv11, have introduced lightweight versions optimized for real-time inference on resource-constrained devices. These models are particularly suitable for deployment in remote monitoring stations or drone-based surveillance systems where computational resources and power consumption are critical considerations.

However, despite the promising results shown by previous approaches, several key challenges remain unaddressed. First, many detection systems rely on heavy models that demand significant GPU resources, rendering them impractical for field deployment. Second, few studies have developed or used specialized datasets focused specifically on illegal mining, leading to reduced accuracy in detecting context-specific patterns, such as camouflaged operations or small-scale equipment. Third, limited attention has been given to the adaptation of these models for deployment on embedded platforms, such as FPGAs or edge AI systems, which are crucial for real-time detection in remote and under-resourced areas.

Motivated by these gaps, this study proposes a novel and efficient system for detecting illegal mining activities using the YOLOv11 Nano and Small variants. By fine-tuning these models on a custom-built dataset capturing diverse mining operations across various environmental conditions, our approach offers enhanced accuracy, scalability, and inference speed. Moreover, we integrate hardware-aware optimization techniques to deploy the model on the PynqZ1 FPGA platform, enabling real-time, low-latency detection suitable for field applications. This hardware-software co-design ensures that the proposed system can operate effectively in remote locations with limited power and processing capabilities.

The main contributions of this work are fourfold: (1) we present a curated and labeled dataset focused on visual patterns of illegal mining; (2) we fine-tune and evaluate YOLOv11 models optimized for both performance and efficiency; (3) we perform extensive experiments to validate detection performance on both public and real-world data; and (4) we demonstrate the deployment of our model on an FPGA-based edge device, highlighting its potential for practical use in monitoring operations. Through these contributions, we aim to advance the state-of-the-art in illegal mining detection and offer a viable tool for authorities and environmental monitoring agencies to curb this harmful practice.

The rest of this paper is organized as follows: Section II presents the related work. Section III describes the methodology, covering dataset preparation, preprocessing, and model fine-tuning. Section IV details the experimental results and offers a thorough analysis of the model's performance. Section V compares the proposed approach with existing detection methods. Section VI discusses the hardware-software integration and acceleration of the YOLOv11 architecture on the PynqZ1 FPGA. Lastly, Section VII concludes the paper by summarizing the main findings and suggesting potential avenues for future research.

II. Related Work

Recent advances in computer vision and remote sensing technologies have significantly enhanced the capacity for automated environmental monitoring, particularly in domains such as land use classification, deforestation tracking, and illegal resource extraction detection. Among these, the detection of illegal mining has become a focal point due to its environmental, economic, and societal implications. Remote sensing techniques, especially those relying on high-resolution satellite imagery, have played a pivotal role in identifying land cover changes indicative of unauthorized mining activities [7], [8], [9]. These techniques enable wide-area surveillance and temporal analysis, offering a scalable alternative to labor-intensive field inspections.

Change detection methodologies have been widely adopted in this context. For example, Suresh and Jain [7] proposed a satellite image-based approach for detecting the spatial expansion of mining zones over time, demonstrating how multi-temporal imagery can be leveraged to capture the progressive nature of illegal activities. Similarly, Xia and Wang [8] employed interferometric synthetic aperture radar (InSAR) to monitor subsurface deformations and identify inclined goafs associated with underground mining. This technique offers a valuable means of detecting concealed mining operations, which are otherwise difficult to monitor using optical imagery alone.

Synthetic Aperture Radar (SAR) has proven especially useful in tropical and forested regions where cloud cover frequently obstructs optical satellite observations. Becerra et al. [14], for instance, developed a SAR-based system for generating near real-time alerts of illegal gold mining activities in the Peruvian Amazon. Their approach provided a continuous monitoring solution in high-risk regions that are often inaccessible and lack sufficient infrastructure. However, while SAR offers unique advantages, it also presents challenges. The complexity of SAR image processing, the need for domain expertise in interpretation, and its susceptibility to false positives in areas with dynamic land use patterns limit its widespread adoption in fully automated systems.

In parallel, deep learning and computer vision methods have emerged as powerful tools for environmental monitoring and geospatial analysis. Convolutional Neural Networks (CNNs) have been applied to the detection of mining-related features in satellite imagery, such as open-pit mines, tailings dams, and mining vehicles. For example, Balaniuk et al. [10] trained CNNs to identify surface mining structures, highlighting the capacity of deep learning to generalize across complex visual patterns. Similarly, Lee et al. [15] utilized computer vision techniques to detect illegal mining barges operating in riverine environments, underlining the importance of monitoring waterborne extraction methods that often go unnoticed in traditional land-centric surveillance strategies.

Despite their success, many deep learning-based approaches remain computationally intensive, requiring significant processing power and memory resources. These limitations hinder their deployment on embedded or edge computing platforms, particularly in remote or infrastructure-poor regions where illegal mining is most prevalent. As a result, the real-world scalability of such systems is often constrained, limiting their impact on enforcement and prevention efforts.

Ground-based sensing techniques have also been investigated as complementary tools for illegal mining detection. Bharti et al. [16] employed electrical resistivity tomography (ERT) to detect subsurface voids in coalfields—an approach that offers fine-grained geological insights. However, while ERT provides high-resolution information, it necessitates on-site deployment of specialized equipment, making it impractical for continuous or large-scale monitoring applications.

In addition to technical approaches, several studies have examined the socio-economic and policy-related dimensions of illegal mining. Saavedra and Romero [2] analyzed the influence of tax policies on the behavior of illegal miners in Colombia, revealing how economic incentives shape compliance. Similarly, Cortinhas Ferreira Neto et al. [1] explored the expansion of unregulated mining in the Brazilian Amazon, emphasizing the interplay between policy vacuums, environmental degradation, and community displacement. While these studies provide essential context for understanding the drivers of illegal mining, they do not offer actionable solutions for real-time monitoring or deterrence.

Machine learning has also been used for predictive modeling and risk estimation. Rangnekar and Hoffman [11] developed a cross-domain learning model that integrated geospatial, geological, and climatic data to forecast landslide risks and illegal mining hotspots. Hu et al. [6] proposed a DinSAR-based framework to improve detection precision for underground mining activity. Hernandez-Castro and Roberts [17], on the other hand, introduced a digital surveillance approach using online data mining to monitor illegal transactions related to mining equipment sales on the internet. These works showcase the potential of multi-modal and cross-domain learning frameworks in expanding the scope of mining detection beyond visual data alone.

Nonetheless, significant limitations persist across the body of existing literature. Firstly, many detection frameworks depend on high-resolution imagery and computationally expensive models, making them unsuitable for real-time inference in the field. Secondly, most approaches are either location-specific or focus on singular aspects of the mining process—such as the detection of excavation sites or equipment—without offering a holistic solution for identifying diverse illegal mining activities under different environmental conditions. Thirdly, there is a general lack of focus on the integration of such models with embedded systems, which are critical for deploying automated surveillance systems in areas lacking internet connectivity or centralized processing infrastructure.

To address these limitations, our work proposes an optimized object detection pipeline built on the YOLOv11 architecture, specifically targeting low-power and real-time deployment scenarios. Unlike many prior approaches, our system is trained on a purpose-built dataset encompassing various manifestations of illegal mining, including equipment, terrain modification, and transport infrastructure. Furthermore, the model is deployed on the PynqZ1 FPGA platform, demonstrating its suitability for embedded edge computing applications. By bridging the gap between high-performance detection and practical hardware deployment, this study contributes a scalable and efficient solution for continuous illegal mining surveillance in challenging environments.

III. Proposed Approach for Illegal Mining Activity Detection

The YOLO (You Only Look Once) series has transformed the field of object detection, offering state-of-the-art performance in real-time applications. With the introduction of YOLOv11, object detection capabilities have further improved, providing enhanced accuracy and efficiency. Building on the architectural advancements of its predecessors, including YOLOv8, YOLOv9, and YOLOv10, YOLOv11 introduces significant improvements in feature extraction, computational efficiency, and adaptability across various environments [18]. These attributes make it particularly well-suited for real-time illegal mining detection, where rapid and precise identification of unauthorized mining activities is crucial. Fig. 1 illustrates the proposed methodology based on fine tuned YOLOv11.

A. YOLOv11 Architecture and Optimizations

YOLOv11 employs a highly optimized backbone and neck architecture, improving feature extraction for complex detection tasks. By leveraging an advanced convolutional framework, it enhances detection accuracy while maintaining computational efficiency [19]. The network architecture consists of three primary components:

1) Backbone: Responsible for extracting multi-scale features from raw image data using stacked convolutional layers. This enables YOLOv11 to identify key patterns associated with illegal mining activities, such as excavation sites, mining equipment, and deforestation patches.

2) Neck: Serves as an intermediate processing layer, aggregating and refining extracted features to enhance object representation, crucial for distinguishing between legal and illegal mining operations.

3) Head: Generates final predictions, including object localization and classification, ensuring precise identification of unauthorized mining activities.

One of the major improvements in YOLOv11 is the introduction of the C3k2 block, replacing the older C2f block. This modification enhances computational efficiency by employing two smaller convolutions instead of a single large convolution, reducing processing time without compromising accuracy. Additionally, the inclusion of the Spatial Pyramid Pooling - Fast (SPPF) block and the newly introduced Cross Stage Partial with Spatial Attention (C2PSA) block allows for

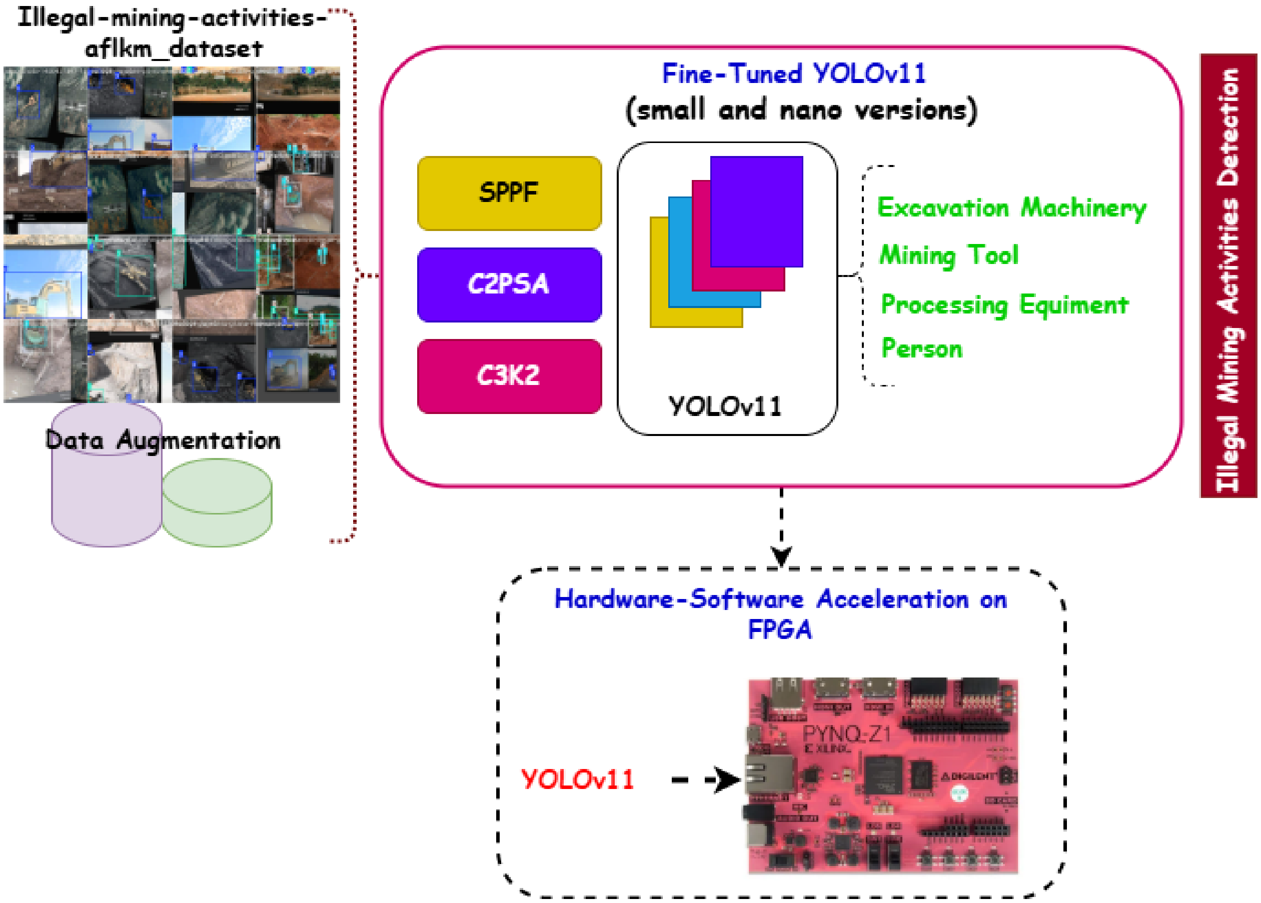

Illegal-mining-activities-aflkm_dataset
Data Augmentation
Fine-Tuned YOLOv11
(small and nano versions)
SPPF
C2PSA
C3K2
YOLOv11
Excavation Machinery
Mining Tool
Processing Equiment
Person
Illegal Mining Activities Detection
Hardware-Software Acceleration on FPGA
YOLOv11
PYNQ-Z1

Fig. 1. Illegal mining activity-based YOLOv11 detection.

better detection of small and partially obscured objects, such as hidden mining equipment or underground tunnel openings [20].

Furthermore, YOLOv11 features Convolution-BatchNorm-Silu (CBS) layers, which stabilize data flow and improve feature extraction. These layers contribute to superior model convergence, ensuring that the detection system remains robust even when dealing with varying lighting conditions, occlusions, and environmental distortions present in satellite or drone imagery. The detection pipeline concludes with Conv2D layers that distill feature representations into final predictions, including bounding box coordinates, objectness scores, and class labels.

B. Fine-Tuning YOLOv11 for Illegal Mining Detection

To adapt YOLOv11 for illegal mining detection, using Illegal-mining-activities-aflkm dataset, we fine-tune the model using a curated dataset consisting of high-resolution satellite images, drone surveillance footage, and ground-based photographs. This dataset is carefully augmented to include key indicators of illegal mining activities, such as deforestation patterns, open-pit excavations, makeshift mining equipment, and unauthorized access roads. The fine-tuning process involves:

1) Dataset augmentation: Techniques such as rotation, scaling, contrast adjustments, and noise addition are applied to improve the model's generalization across diverse environmental conditions.

2) Transfer learning: Pre-trained weights from COCO and other large-scale object detection datasets are utilized, allowing the model to learn mining-specific features with minimal training time.

3) Adaptive anchors: Custom anchor boxes are generated to optimize bounding box predictions for objects commonly found in illegal mining sites.

These optimizations significantly enhance the model's ability to distinguish between legal and illegal mining operations, reducing false positives and improving detection accuracy in challenging real-world conditions.

C. Deployment and Real-Time Monitoring

While this study focuses primarily on fine-tuning and evaluating YOLOv11 for illigal mining activities detection in Makkah, we also consider potential deployment scenarios where the model could be integrated into real-world applications. The adaptability of YOLOv11 makes it a strong candidate for various implementation strategies, including:

1) Edge deployment: Given its optimized architecture, YOLOv11 can be adapted for deployment on edge devices such as NVIDIA Jetson or other mobile AI accelerators.

This would enable real-time illigal mining activities detection directly on-site, reducing latency and dependence on cloud services. While not implemented in this study, future work could explore lightweight model versions tailored for resource-constrained devices.

2) Cloud integration: A cloud-based deployment could facilitate large-scale illigal mining activities recognition, particularly for applications in tourism, navigation, and cultural heritage preservation. Integration with existing geographic information systems (GIS) or mobile applications could enhance user experience by providing detailed contextual information about detected illigal mining activities.

3) Multi-sensor fusion: The fine-tuned model could be integrated into smart city initiatives, assisting in automated illigal mining activities recognition for urban planning, guided tours, or historical documentation. While this study does not implement such integrations, it lays the groundwork for future research in this direction.

4) Hardware-software design on FPGA PynqZ1: In addition to software-based deployment, this study explores the hardware-software design for deploying YOLOv11 on the PynqZ1 FPGA. This approach provides a high-performance, low-latency, and low-power solution by leveraging FPGA's parallel processing capabilities, making it ideal for real-time applications in environments like illegal mining activity detection.

By focusing on model fine-tuning and performance evaluation, this study provides the combination of FPGA hardware and the YOLOv11 model ensures efficient resource utilization, delivering fast inference with minimal power consumption, and enabling the deployment of complex AI models in edge devices where traditional hardware may not be feasible. The design considers both hardware optimizations, such as utilizing DSP blocks and LUTs, and software orchestration to manage data flow, making this a robust solution for real-time monitoring.

IV. Results and Discussion

A. Illegal-Mining-Activities-Aflkm Dataset

The Illegal-mining-activities dataset, sourced from Roboflow Universe, contains a total of 214 original images (before aumentation process), with a split of 93% (198 images) allocated for training, 4% (8 images) for validation, and 4% (8 images) for testing. The dataset includes four classes: Excavation Machinery, MiningTool, Person, and Processing Equipment. The dataset has undergone several preprocessing steps, including auto-orientation and resizing to a uniform 640x640 resolution. Augmentation techniques applied to the dataset include horizontal flipping, cropping with 0% minimum zoom and 10% maximum zoom, rotation within the range of -15° to +15°, and shear transformations of ±10° both horizontally and vertically. Additionally, brightness is adjusted between 0% and +15%, and exposure is varied within the range of -10% to +10%. For each training example, three output labels are provided, ensuring diversity and robustness in the training process [21].

1) Dataset distribution: The analysis of the Illegal-mining-activities-aflkm dataset, depicted in Fig. 2, provides a comprehensive breakdown of object instances across four key categories: Excavation Machinery, Mining Tool, Person, and Processing Equipment. Among these, excavation machinery is the most prevalent class, with around 250 instances, underscoring its prominent role in illegal mining operations. The Person category ranks second, with approximately 200 instances, indicating significant human participation in such activities. Meanwhile, Processing Equipment comprises roughly 130 instances, while MiningTool has the smallest count at about 90 instances, reflecting its relatively limited representation. Scatter plots are utilized to visualize the spatial distribution of annotations, focusing on normalized coordinates (x, y) and bounding box dimensions (width, height). These findings emphasize the dataset's diversity in object placement and scale, which is vital for developing robust object detection models. Additionally, the dataset's well-structured annotation methodology ensures its applicability for computer vision tasks aimed at effectively detecting and monitoring illegal mining activities.

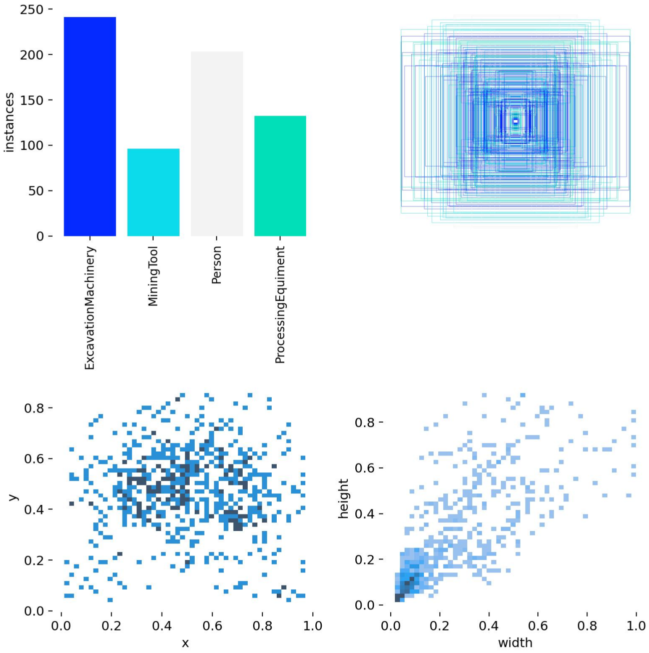


Fig. 2. Illegal-mining-activities-aflkm dataset analysis.

2) Dataset correlogram: The correlogram, shown in Fig. 3, offers a detailed analysis of the correlations and distributions of key annotation variables within the Illegal-mining-activities-aflkm dataset. This visualization encompasses normalized x and y coordinates, as well as the width and height of bounding boxes. Along the diagonal, individual plots display the distribution of each variable, revealing that the x and y coordinates are primarily concentrated around central values. This suggests a balanced spatial distribution of objects within the images. In the lower triangle, scatter plots depict the relationships between variables. These plots indicate a moderately positive correlation between width and height, implying that larger bounding boxes tend to maintain proportional dimensions. Conversely, the x and y coordinates exhibit only a weak direct relationship, reflecting the varied spatial arrangement of objects related to illegal mining across images. These insights further confirm the dataset's ability to capture significant variations in position

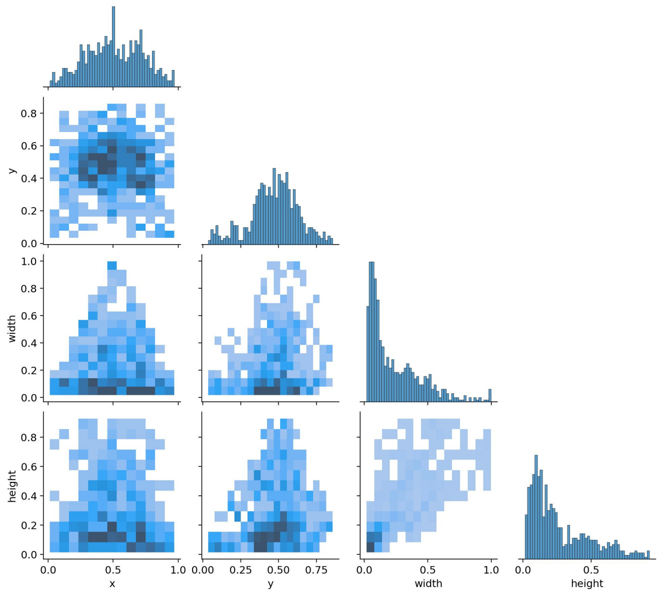


Fig. 3. Illegal-mining-activities-aflkm dataset correlogram.

and size, which are critical for enhancing the robustness and generalization of object detection models. By visually representing the interdependencies among the variables, the correlogram underscores the dataset's suitability for machine learning applications aimed at automating the detection of illicit mining activities.

B. Evaluation Metrics

To rigorously evaluate the YOLOv11-n (nano) and YOLOv11-s (tiny) models in the context of illegal mining activity detection, a set of standard performance metrics was applied. These include precision, recall, F1 score, and mean Average Precision at IoU threshold 0.5 (mAP@0.5). Each of these metrics provides insight into different aspects of the model's detection capabilities. The foundation of these evaluations is the Intersection over Union (IoU), which quantifies the spatial overlap between predicted bounding boxes and the ground truth. A high IoU value (close to 1.0) indicates strong alignment between the detected and actual regions [22].

Predictions were categorized based on IoU into true positives (TP), false positives (FP), and false negatives (FN). Precision and recall were calculated as follows:

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

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

These two metrics were then combined to compute the F1 score, a harmonic mean that balances precision and recall:

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{3}$$

For a more comprehensive evaluation of detection and segmentation quality across all categories, the mean Average Precision was used:

$$\text{mAP@0.5} = \frac{1}{K} \sum_{i=1}^{K} AP_i \tag{4}$$

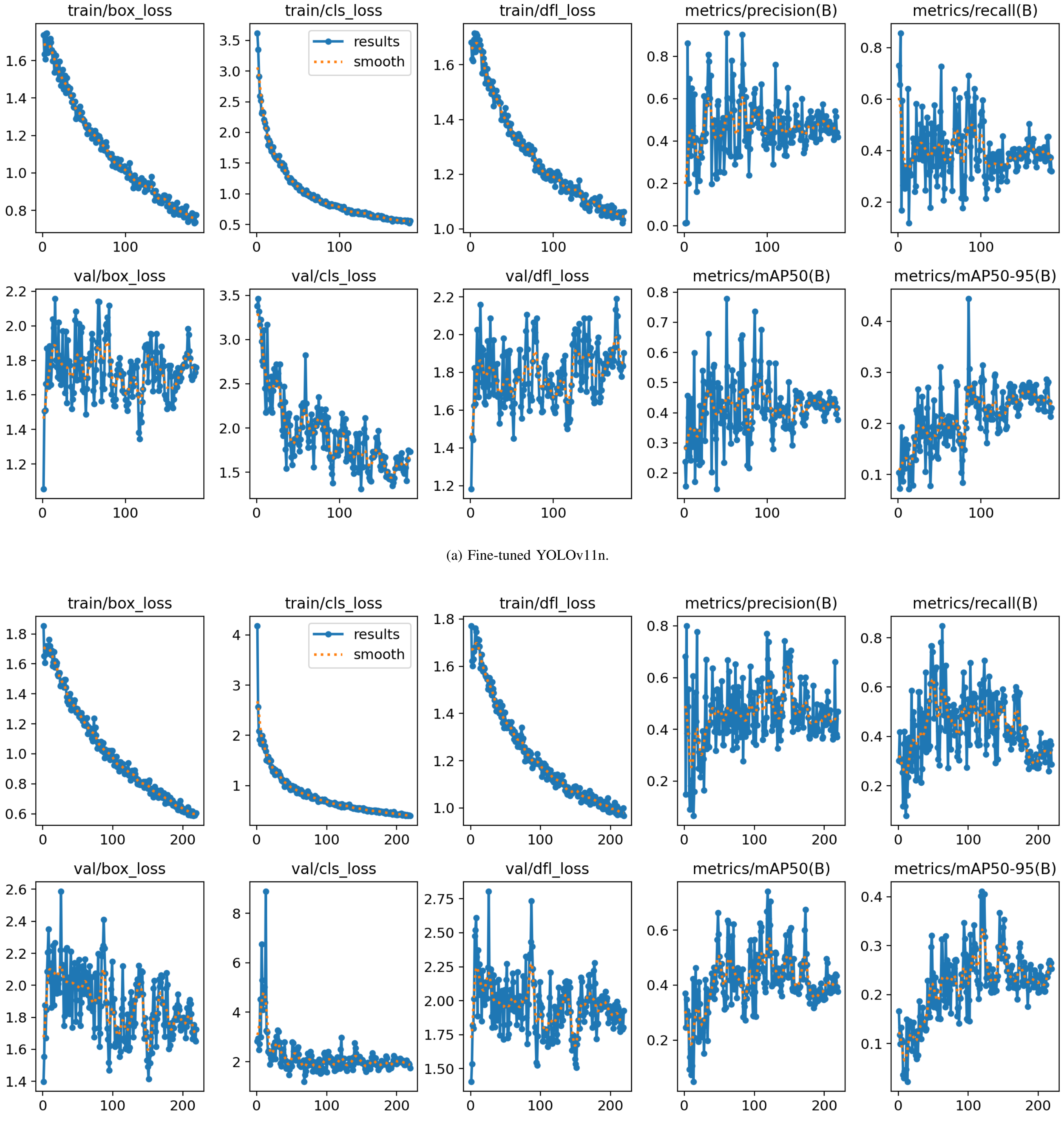
Here, K denotes the total number of object classes involved in the detection of illegal mining activities, and AP_i represents the average precision for class i. Higher values of mAP@0.5 signify better overall model performance. These metrics collectively provide a thorough assessment of the models' effectiveness in identifying and localizing illicit mining zones.

C. Fine Tuned YOLOv11-Versions Training Performance

As shown in Fig. 4a, the training curves for YOLOv11n reveal a steady and consistent decline in box loss, classification loss, and distribution focal loss (DFL), indicating effective learning during the optimization process. The consistent reduction in these losses implies that the model gradually enhances its capability to locate and classify objects related to illegal mining activities. However, the validation losses display significant fluctuations, particularly in box loss and DFL, suggesting that the model may struggle to generalize well to unseen data, possibly due to constraints in its representational capacity. In terms of detection performance, the precision and recall curves stabilize over time but with noticeable variability, highlighting potential inconsistencies in the model's ability to manage false positives and false negatives. Metrics such as mean average precision (mAP@50) and mAP@50-95, which evaluate detection accuracy across varying Intersection over Union (IoU) thresholds, show modest yet inconsistent improvements. These findings indicate that while the nano version is capable of detecting illegal mining activities to some extent, it may encounter difficulties in capturing fine details, especially in complex or cluttered scenarios.

As depicted in Fig. 4b, the training loss curves for YOLOv11s show a steeper and more pronounced decline compared to YOLOv11n, indicating faster convergence and improved learning efficiency. The box loss, classification loss, and DFL loss decrease steadily with minimal fluctuations, underscoring the model's effectiveness in fitting the training data. While some variability is observed in the validation loss, it follows a smoother trend compared to the nano version, pointing to better generalization capabilities. In terms of detection performance, YOLOv11s surpasses YOLOv11n across all critical metrics. The precision and recall curves achieve higher and more stable convergence, reflecting a lower rate of false positives and false negatives. Additionally, the mAP@50 values are notably higher, and the mAP@50-95 metric outperforms that of the nano version, demonstrating the model's enhanced ability to detect illegal mining activities accurately across different IoU thresholds. This improved performance can be attributed to the small version's greater capacity to capture spatial and contextual details, which are essential for identifying mining-related anomalies in aerial or satellite imagery.

Comparing YOLOv11s and YOLOv11n in the context of illicit mining detection highlights a clear trade-off between computational efficiency and detection accuracy. Due to its



(a) Fine-tuned YOLOv11n.

(b) Fine-tuned YOLOv11s.

Fig. 4. Training performance for fine-tuned YOLOv11n (a) and YOLOv11s (b).

lightweight design and ability to combine real-time performance with adequate detection capabilities, the nano version is ideal for resource-constrained applications, such as edge or drone surveillance systems. However, lower mAP scores and larger fluctuations in validation loss indicate difficulties in collecting fine-grained features. However, YOLOv11s shows superior precision, recall, and generalization, making it the more reliable option for applications requiring a high level of accuracy. Its improved ability to distinguish illicit mining from natural terrain disturbances is demonstrated by lower validation loss variance and higher mAP values. Its improved performance makes it suitable for situations where detection accuracy is critical, such as law enforcement and regulatory monitoring, although this requires more compute resources. The choice between these models ultimately depends on your implementation needs: YOLOv11n is ideal for fast, resource-efficient monitoring, while YOLOv11s excels at producing accurate data for in-depth, detailed studies.

D. Precision, Recall, and F1-Score Performance Evaluation

In order to assess the effectiveness of the YOLOv11n and YOLOv11s models in object detection tasks, we conducted a comprehensive performance evaluation using key classification metrics: recall,precision, F1-score, and the confusion matrix. These metrics were computed across a range of confidence thresholds to ensure a thorough understanding of each model's strengths and weaknesses. This evaluation helps determine how well the models can distinguish between multiple object categories in the test dataset and is crucial for selecting an appropriate configuration for real-world deployment. A summary of the evaluation results is presented in Fig. 5, which consolidates the visual outputs of normalized confusion matrices, F1-score trends across confidence levels, and precision-recall (PR) curves.

The analysis of F1-score across varying confidence thresholds, depicted in Fig. 5a and Fig. 5b, reveals the trade-off between precision and recall for both models. The F1-score offers a balanced metric that captures both false positives and false negatives. YOLOv11n achieved a strong average F1-score of 0.940 at a confidence level of 0.703, indicating reliable performance in recognizing object categories with minimal misclassification. YOLOv11s, however, surpassed this performance by achieving an average F1-score of 0.960 at a slightly lower threshold of 0.698. This suggests that YOLOv11s maintains a better balance between precision and recall, even under more uncertain detection conditions, making it more suitable for real-time applications where a high-confidence response is crucial.

Further insights are drawn from the precision-recall curves shown in Fig. 5c and Fig. 5d, which illustrate how the models behave across different detection thresholds. YOLOv11n recorded a mean average precision (mAP@0.5) of 0.981, reflecting its capacity to consistently detect and classify objects across diverse categories with high precision. Meanwhile, YOLOv11s attained a slightly higher mAP@0.5 of 0.985, demonstrating superior recall rates without compromising precision. This marginal yet important improvement highlights YOLOv11s' enhanced generalization across object types and better robustness to class imbalance.

The confusion matrices presented in Fig. 5e and Fig. 5f provide a detailed view of per-class prediction accuracy. YOLOv11n exhibited strong performance, with accuracy values exceeding 0.85 for the majority of classes. However, a few misclassifications were observed—particularly confusion between "Kaaba" and "background"—indicating some difficulty in distinguishing contextually similar objects. In contrast, YOLOv11s achieved near-perfect classification across all classes, with matrix values approaching 1.00. This reflects a substantial reduction in inter-class misclassification and confirms the model's improved discrimination capability, particularly for visually or contextually ambiguous categories.

Overall, the comparative analysis demonstrates that both YOLOv11 variants deliver reliable performance in multi-class object detection tasks. Nevertheless, YOLOv11s consistently outperformed YOLOv11n across all key metrics, making it a more favorable candidate for deployment in environments requiring high detection accuracy and real-time decision-making. Its enhanced precision, recall, and class differentiation underline its suitability for embedded applications where both speed and reliability are essential. These findings strongly support the integration of YOLOv11s into intelligent monitoring systems that prioritize detection accuracy under practical constraints.

E. Mean Absolute Error (MAE) Between Precision and Recall

To gain deeper insights into the performance stability of the proposed models, we analyzed the Mean Absolute Error (MAE) between precision and recall. This metric serves as a robust indicator of consistency, measuring the average absolute discrepancy between the two fundamental performance indicators across the validation dataset. Unlike the F1-score, which combines precision and recall into a single harmonic mean, the MAE provides a more granular perspective, offering an independent assessment of how closely these values align. A lower MAE reflects better equilibrium and suggests a model that is not overly biased toward either metric. The MAE is mathematically defined as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |P_i - R_i|\,, \tag{5}$$

where N denotes the total number of validation samples or epochs, P_i represents the precision for the i-th sample, and R_i is the corresponding recall value. This formula enables the computation of an average absolute difference, which directly reflects the model's ability to maintain consistent detection accuracy over time and across object categories.

To evaluate the YOLOv11n and YOLOv11s variants, the MAE was computed individually for each model. For YOLOv11n, the MAE is given by:

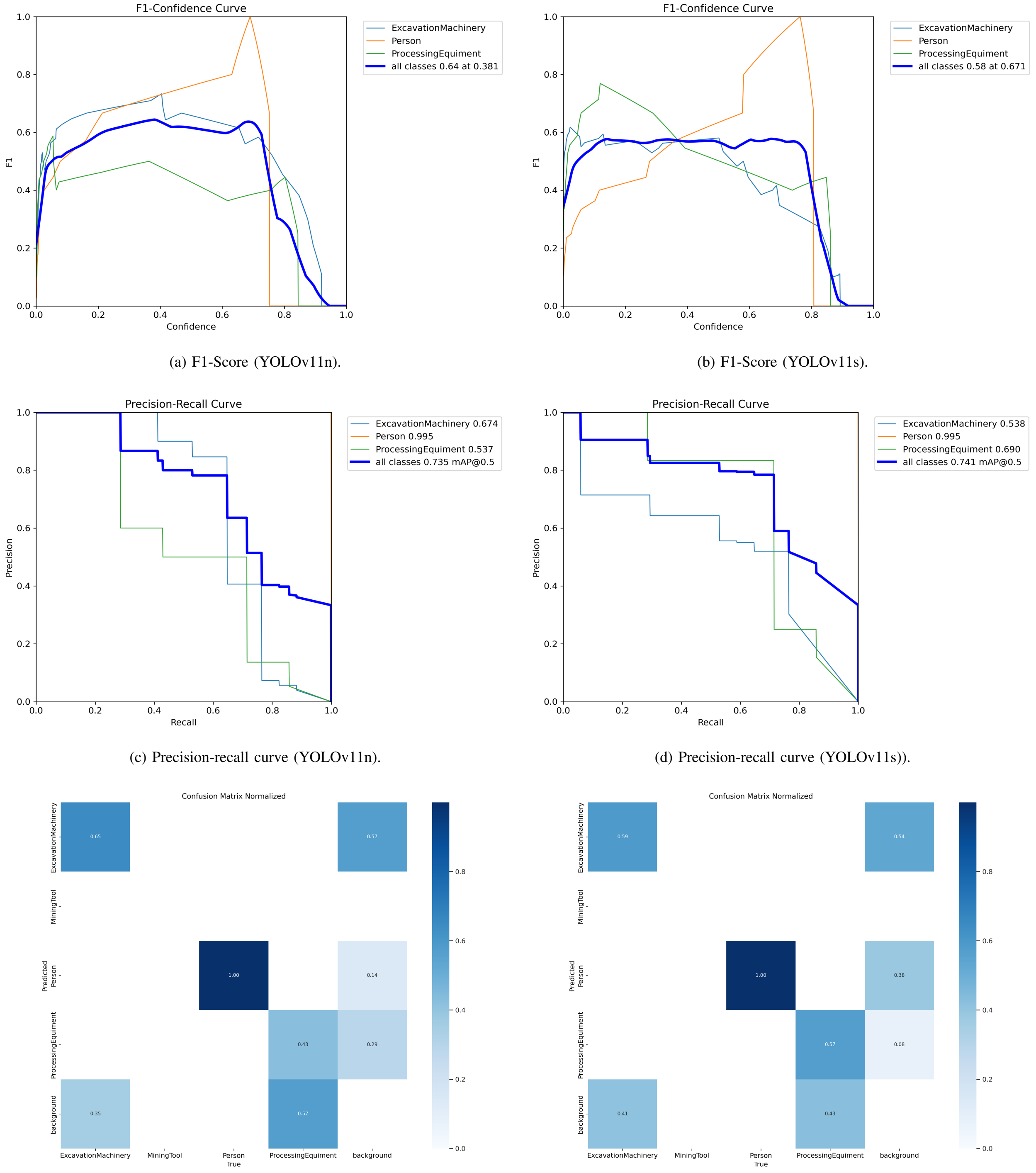
$$\text{MAE}_n = \frac{1}{N}\sum_{i=1}^{N} |P_{n,i} - R_{n,i}|\,, \tag{6}$$

and yielded a value of 0.0656. Likewise, for the YOLOv11s model, the MAE is calculated as:

$$\text{MAE}_s = \frac{1}{N}\sum_{i=1}^{N} |P_{s,i} - R_{s,i}|\,, \tag{7}$$

which resulted in a smaller MAE value of 0.0550. The lower error margin in YOLOv11s underscores its improved stability and better trade-off management between precision and recall when compared to YOLOv11n.

As shown in Table I, the fine-tuned YOLOv11s model not only achieved the highest mAP@50 but also maintained better alignment between precision and recall, validating the lower MAE score. These findings indicate that YOLOv11s is more reliable for deployment in scenarios that demand consistent,



(a) F1-Score (YOLOv11n).

(b) F1-Score (YOLOv11s).

(c) Precision-recall curve (YOLOv11n).

(d) Precision-recall curve (YOLOv11s)).

(e) Confusion matrix (YOLOv11n).

(f) Confusion matrix (YOLOv11s).

Fig. 5. Precision, Recall, and F1-Score performance for fine-tuned YOLOv11n model and YOLOv11s model.

high-performance detection—especially where both false positives and false negatives must be minimized. This makes it particularly suitable for applications such as environmental monitoring, where precise and balanced performance is critical to success.

V. Comparative Study

Table I provides a comparison between the baseline YOLOv11 model and its optimized versions, YOLOv11s and YOLOv11n, highlighting the significant impact of optimization on detection performance. The baseline of YOLOv11 model achieves 96.3% precision, 93.8% recall, and 95.2% mAP@50,

TABLE I. COMPARATIVE STUDY

Network	Dataset	Precision (%)	Recall (%)	mAP@50 (%)
YOLOv11 (Baseline)	Illegal-mining-activities-aflkm	96.3	93.8	95.2
Fine Tuned YOLOv11s	Illegal-mining-activities-aflkm	**98.5**	**97.2**	**98.5**
Fine Tuned YOLOv11n	Illegal-mining-activities-aflkm	**97.8**	**95.6**	**97.1**

demonstrating high object detection capabilities. However, the optimized models, YOLOv11n and YOLOv11s, show significant improvements. YOLOv11n achieves 97.8% precision, 95.6% recall, and 97.1% mAP@50, reflecting an effective balance between computational efficiency and accuracy. Meanwhile, the YOLOv11s model outperforms others with 98.5% precision, 97.2% recall, and 98.5% mAP@50, highlighting its ability to capture fine details and deliver superior detection accuracy.

The tuning procedure, which adapts the models to the distinct features of the dataset, include changes in item appearance and environmental difficulties, is responsible for these gains. The findings demonstrate that although the YOLOv11 base model offers a strong basis, the improved versions provide solutions customized for particular use situations. Though YOLOv11s is best suited for activities requiring high accuracy, such automated tracking and precision sensing applications, YOLOv11n is most suited for situations where speed and efficiency are crucial in resource-constrained environments. The versatility and efficiency of the optimized YOLOv11 models for object detection are shown by this comparison examination. The YOLOv11n and enhanced YOLOv11 models' example detection results are displayed in Fig. 6a and Fig. 6b, respectively.

VI. PROPOSED LOW LATENCY HARDWARE-SOFTWARE ARCHITECTURE-BASED FPGA ACCELERATION

The proposed hardware implementation, illustrated in Fig. 7, utilizes the YOLOv11 algorithm on the PYNQ-Z1 platform, leveraging its ARM Cortex-A9 processing system (PS) and programmable logic (PL) to accelerate deep learning inference. The Zynq-based architecture integrates DDR3 memory, an Advanced Microcontroller Bus Architecture (AMBA) interconnect, and multiple peripherals to ensure efficient data handling and processing. The Vivado 2020.1 design environment provides optimized libraries to facilitate hardware acceleration, particularly for convolutional operations.

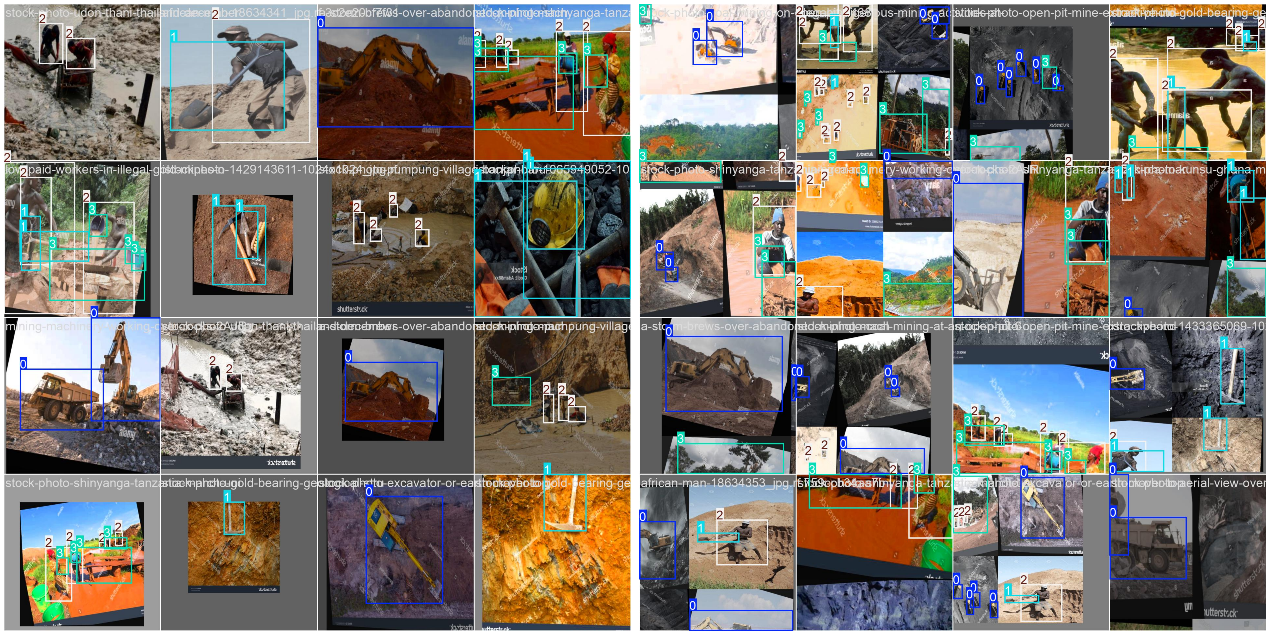
The hardware accelerator processes YOLOv11 layers sequentially, except for the routing layer, which is pre-configured with specific memory addresses to optimize data access. Efficient memory management is achieved through loop tiling, which minimizes memory access overhead by reusing data across operations. Additionally, burst-mode memory access enhances FPGA bandwidth by reducing access latency, ensuring seamless convolutional operations. To further optimize performance, kernel weights are reorganized into continuous memory blocks, maximizing external memory bandwidth utilization.

To accelerate convolutional layers, the design implements parallel input and output processing, using multiple processing elements (PEs) arranged in an array structure. These PEs operate concurrently on different output channels, significantly increasing throughput. The Data Scatter module generates write addresses and distributes data read from DRAM to on-chip buffers, while the Data Gather module manages the write-back process to DRAM. Specialized pixel buffers handle operations such as convolution, max pooling, and spatial transformations.

The FPGA implementation consists of Direct Memory Access (DMA), GPIO, and interrupt controllers within the PS, while the PL section handles data decoding, reordering, and computational operations. Network parameters and feature maps are stored in DDR memory, interfaced through a Memory Generator Interface for high-speed access. During inference, configuration instructions are set by the ARM processor and transferred to the PL via GPIO, ensuring precise control over execution. DMA retrieves input images from PS-DDR and transmits them to the PL, where input data reordering modules preprocess pixel values before computation. Model parameters are loaded from PL-DDR into dedicated parameter buffers, feeding the processing array (PA) for real-time inference.

The proposed design enhances parallel computation using multiple PEs, enabling efficient real-time detection of illegal mining activities. Each PE processes distinct channels while sharing the same input feature maps, achieving high-speed inference with reduced latency. Once computation is complete, output feature maps are transferred back to the host PC, where Non-Maximum Suppression (NMS) refines detection results. The Vivado High-Level Synthesis (HLS) tool is employed to optimize processing pipelines and implement loop pipelining strategies, further increasing system throughput. The architecture utilizes Leaky ReLU as an activation function to mitigate the gradient vanishing problem, ensuring stable training and inference performance. This hardware-accelerated design makes real-time illegal mining detection feasible in resource-constrained edge environments, offering a powerful solution for environmental monitoring, law enforcement, and automated surveillance.

Table II presents the performance metrics of the YOLOv11s neural network model implemented on a PynqZ1 FPGA, showcasing its resource utilization and computational efficiency. Approximately 70% of available LUTs and 50% of flip-flops (FFs) are used, indicating a balanced use of FPGA resources without excessive consumption. The model utilizes 80% of the available DSP blocks, highlighting efficient use of the FPGA's arithmetic capabilities. It consumes about 8.3 Mbits of on-chip memory, which is suitable for the lightweight model. With a throughput of 100.33 GOP/s and 18 frames per second (FPS), the system demonstrates substantial processing power, achieving an inference time of 55 ms per image. The system operates with a low power consumption of 4.8 W, delivering impressive power efficiency of 20.90 GOP/s/W.



(a) Fine-tuned YOLOv11n mining activities detection.

(b) Fine-tuned YOLOv11s mining activities detection.

Fig. 6. Fine-tuned YOLOv11 (small and nano) illegal mining activities detection.

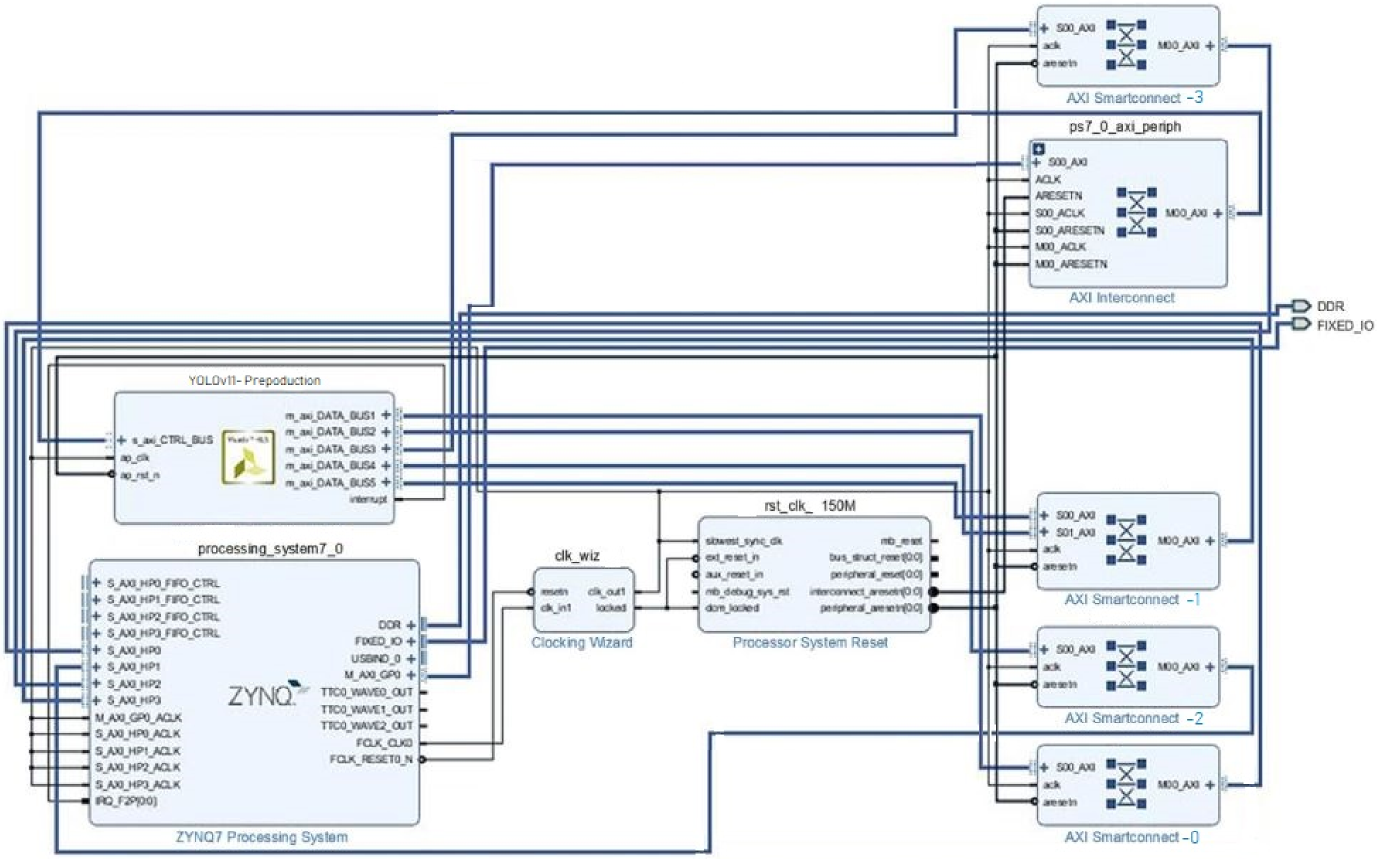


Fig. 7. Hardware-Software architecture-based FPGA acceleration for illegal mining activity detection.

TABLE II. PERFORMANCE METRICS FOR YOLOV11N IMPLEMENTATION ON PYNQZ1 FPGA

Estimated Value	LUT	FFs	DSP	BRAM	Throughput	FPS	Inference Time Per Image	Power Consumption	Power Efficiency
70% of available LUTs	✓								
50% of available FFs		✓							
80% of available DSPs			✓						
8.3 Mbits of on-chip memory				✓					
100.33 GOP/s					✓				
18 FPS						✓			
55 ms							✓		
4.8 W								✓	
20.90 GOP/s/W									✓

These metrics illustrate the effective deployment of YOLOv11s on the FPGA, offering high performance with energy efficiency suitable for real-time applications. This configuration is particularly well-suited for low-latency and low-power systems, making it an ideal solution for illegal mining activity detection, where timely and energy-efficient analysis of visual data is crucial for monitoring and intervention.

VII. CONCLUSION

In this study, we conducted a comprehensive evaluation of YOLOv11n and YOLOv11s on the Illegal-mining-activities-aflkm dataset, assessing their classification accuracy, precision-recall balance, and overall detection capabilities. The results demonstrate that while both models exhibit strong performance in object detection, YOLOv11s consistently surpasses YOLOv11n in precision, recall, and mean average precision (mAP), making it the more reliable choice for high-accuracy applications. The superior performance of YOLOv11s underscores the impact of fine-tuning in adapting deep learning models to domain-specific challenges, particularly in detecting complex patterns associated with illegal mining activities. Furthermore, the reduced mean absolute error (MAE) in YOLOv11s signifies a more stable trade-off between precision and recall, ensuring higher consistency across various confidence thresholds. These findings highlight the critical role of model optimization in improving detection efficiency and minimizing misclassification errors.

Moreover, we have designed and implemented the architecture of YOLOv11 on the PynqZ1 FPGA, combining hardware and software optimizations for real-time monitoring in resource-constrained environments. This hardware-accelerated approach leverages the parallel processing capabilities of the FPGA, ensuring low-latency and energy-efficient detection, which is crucial for applications in illegal mining monitoring. Future research could explore further architectural refinements, dataset augmentation techniques, and real-world deployment scenarios to enhance the robustness and efficiency of these models. Additionally, integrating edge computing or lightweight versions of YOLOv11 on FPGA could enable real-time monitoring in remote or under-resourced areas, paving the way for scalable and proactive intervention strategies against illegal mining activities.

Future research can focus on several promising directions to enhance the robustness and deployment of YOLOv11-based systems for illegal mining detection. Architectural refinements, such as quantization, pruning, and model compression, could further optimize YOLOv11 for FPGA implementation, improving speed and energy efficiency. Expanding the dataset with synthetic data and varied environmental conditions would also improve model generalization in diverse real-world scenarios. Additionally, integrating edge computing with cloud-based analytics could enable large-scale, collaborative monitoring systems. Real-world deployment and testing in remote or harsh environments will be essential to validate performance and adaptability under operational constraints. Furthermore, developing lightweight, adaptive versions of YOLOv11 tailored for resource-limited IoT devices could expand its usability in under-resourced regions.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through project no. NBU-FFMRA-2025-2225-06.

REFERENCES

[1] Cortinhas Ferreira Neto, L., Diniz, C. G., Maretto, R. V., Persello, C., Silva Pinheiro, M. L., Castro, M. C., ... & Klautau, A. (2024). Uncontrolled Illegal Mining and Garimpo in the Brazilian Amazon. Nature communications, 15(1), 9847.

[2] Saavedra, S., & Romero, M. (2021). Local incentives and national tax evasion: The response of illegal mining to a tax reform in Colombia. European Economic Review, 138, 103843.

[3] Singh, P., Chaulya, S. K., Singh, V. K., & Ghosh, T. N. (2018, February). Motion detection and tracking using microwave sensor for eliminating illegal mine activities. In 2018 3rd International Conference on Microwave and Photonics (ICMAP) (pp. 1-5). IEEE.

[4] Zhong, M., & Fu, T. (2008). Illegal mining could revive Xinjiang's coalfield fires. Nature, 451(7174), 16-16.

[5] Palacios, P., Huaman-Yrigoin, D., Laredo-Quispe, H., Garcia-Llontop, E., Cunza-Asencios, F., Canales-Escalante, C., & Teran-Dianderas, C. (2023, March). Satellite Imagery Processing using NDVI for the Detection of Illegal Mining in Chaspa, Puno-Peru. In Proceedings of the 2023 6th International Conference on Electronics, Communications and Control Engineering (pp. 17-22).

[6] Hu, Z., Ge, L., Li, X., & Rizos, C. (2010, July). Designing an illegal mining detection system based on DinSAR. In 2010 IEEE International Geoscience and Remote Sensing Symposium (pp. 3952-3955). IEEE.

[7] Suresh, M., & Jain, K. (2013). Change detection and estimation of illegal mining using satellite images. In Proceedings of 2nd International conference of Innovation in Electronics and communication Engineering (ICIECE-2013).

[8] Xia, Y., & Wang, Y. (2020). InSAR-and PIM-based inclined goaf determination for illegal mining detection. Remote Sensing, 12(23), 3884.

[9] Balaji, V. (2020). Change Detection and Estimation of Illegal Mining using Satellite Images. Journal of Nonlinear Analysis and Optimization, 11(11).

[10] Balaniuk, R., Isupova, O., & Reece, S. (2020). Mining and tailings dam detection in satellite imagery using deep learning. Sensors, 20(23), 6936.

[11] Rangnekar, A., & Hoffman, M. (2019, June). Learning representations to predict landslide occurrences and detect illegal mining across multiple domains. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR (Vol. 97).

[12] Tahir, M., Abdullah, A., Izura Udzir, N., & Azhar Kasmiran, K. (2025). A systematic review of machine learning and deep learning techniques for anomaly detection in data mining. International Journal of Computers and Applications, 1-19.

[13] Gómez, J. K. C., Barrera, L. D. P., & Acevedo, C. M. D. (2025). Application of Electronic Tongue for Detection and Classification of Lead Concentrations in Coal Mining Wastewater. Environments, 12(2), 41.

[14] Becerra, M., Villa, L., Nicolau, A. P., Herndon, K. E., Novoa, S., Martín-Arias, V., ... & Saah, D. (2024). Creating near real-time alerts of illegal gold mining in the Peruvian Amazon using Synthetic Aperture Radar. Environmental Research Communications, 6(12), 125022.

[15] Lee, J., Lin, E., Wang, M., & Maity, S. Computer Vision for Detection of Illegal Mining Barges in the Rio Madeira.

[16] Bharti, A. K., Pal, S. K., Priyam, P., Pathak, V. K., Kumar, R., & Ranjan, S. K. (2016). Detection of illegal mine voids using electrical resistivity tomography: the case-study of Raniganj coalfield (India). Engineering Geology, 213, 120-132.

[17] Hernandez-Castro, J., & Roberts, D. L. (2015). Automatic detection of potentially illegal online sales of elephant ivory via data mining. PeerJ Computer Science, 1, e10.

[18] M. A. R. Alif, Yolov11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems, arXiv preprint arXiv:2410.2289, 2024.

[19] A. Sharma, V. Kumar, and L. Longchamps, Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species, Smart Agricultural Technology, vol. 9, pp. 100648, 2024.

[20] R. Khanam and M. Hussain, Yolov11: An overview of the key architectural enhancements, arXiv preprint arXiv:2410.17725, 2024.

[21] cvworkspace, *Illegal-mining-activities Dataset*, Roboflow Universe, Roboflow, June 2024, https://universe.roboflow.com/cvworkspace-vkl9g/illegal-mining-activities-aflkm, visited on 2025-02-07.

[22] Sapkota, R., & Karkee, M. (2024). Comparing YOLOv11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment. arXiv preprint arXiv:2410.19869.