Road Layer Detection and Volume Calculation Using UAV Technologies and Artificial Intelligence

M. Hammad Iqbal*, Tahir Saleem, M. Junaid Iqbal

Military College of Engineering, National University of Science and Technology, Risalpur, Pakistan

Abstract:

Traditional methods for monitoring road construction often rely on manual measurements and visual inspections, which are time-consuming and prone to errors. This study introduces a novel approach that uses unmanned aerial vehicles (UAVs) and the You Only Look Once (YOLO) model (YOLOv11-seg) to significantly improve the accuracy and efficiency of road layer detection and volume calculation. By utilizing high-resolution images taken by the DJI Phantom 4 Pro drone, the model successfully identifies and segments crucial road layers, such as subgrade, asphalt, and aggregate base. The YOLOv11-seg model demonstrates remarkable classification accuracy, achieving 95% for both asphalt and aggregate base and 83% for subgrade. Additionally, the model reaches a maximum precision of 1.0 at the highest confidence level. Its F1 score is 0.92 at an optimal confidence threshold of 0.779, and it maintains a recall of 0.96 at lower confidence levels. For volumetric analysis, this study evaluates four software tools: Agisoft Metashape, CloudCompare, Civil 3D, and WebODM. It identifies WebODM as the most accurate, achieving volume accuracies of 99.85% for subgrade and asphalt and 99.9% for aggregate base. This application of drone imagery and object detection technology simplifies monitoring and significantly reduces labor and errors associated with traditional road construction monitoring methods.

Keywords:

Road Construction, YOLO, 3D Mapping, Drone technology, Aerial Road Survey

*Corresponding Author:
marwathammad120@gmail.com

1. Introduction

Traditional approaches for identifying layers in road construction and calculating their volumes are often slow and demand considerable labor. Inspectors usually depend on manual tools, like rulers and measuring tapes, to estimate each layer's volume. They also perform visual inspections to monitor construction progress. These traditional techniques mainly use Global Navigation Satellite System (GNSS) receivers and total station instruments [1]. Although total stations can provide accurate measurements through electronic distances and angle calculations, they require a lot of time, effort, and data interpretation. In modern road construction, precise measurements are crucial, but traditional methods often fall short, causing delays and errors. However, until now, no comprehensive software solution is available for calculating the volume of road layers, and no Artificial Intelligence (AI) model has been trained specifically for detecting road layers. Our research aims to fill this gap by proposing a model for detecting road layers and introducing an innovative software-based approach for volume calculation. This solution promises to be more efficient and accurate.

With the rise of new technologies, many operations have seen big improvements. AI and photogrammetry, for instance, have become key players in various fields, especially construction [2]. A significant development is the use of UAVs in the construction sector. UAVs have become an important tool in the construction industry because they are used for improving efficiency and lowering the costs of various construction processes. They're used for tasks like land surveying, creating topographic maps, monitoring construction progress, and calculating stockpile volumes using aerial images [3][4][5]. Drones equipped with Light Detection and Ranging (LiDAR) sensors generate 3D point clouds for roadside object classification [6] and capture detailed aerial images, making volume calculations, such as earthwork volume, more efficient and precise compared to traditional surveying methods [7]. They also facilitate earth volume measurements through 3D modeling, utilizing vertical and oblique drone images for comprehensive comparison measurements [8]. Kim et al. highlight that UAVs provide a reliable alternative to traditional methods such as GNSS, particularly in early road construction stages where accurate volume estimation is crucial [9]. Additionally, new techniques for automatically registering point clouds have been introduced to improve quality control in construction projects [10. These advances show how drone technology can make construction processes more efficient and accurate. However, using drones specifically to calculate the volume of road layers is an area that hasn't been explored much, and this research aims to delve into that.

AI is rapidly advancing and has become a transformative force in the construction industry. It enhances traditional practices, making the construction process more time-efficient and cost-effective. It is used to improve overall efficiency and reduce the need for on-site operators, such as surveyors [11]. AI is currently used for road assessment, including crack and pothole detection. For example, the YOLOv5 object detection algorithm has proven effective in identifying road defects from images captured by UAVs, demonstrating significant potential in construction site monitoring [12]. Furthermore, the combination of convolutional neural networks (CNN) with Markov random fields (MRF) has been shown to improve the accuracy of detecting road defects. This approach helps ensure spatial consistency in feature extraction and classification [13]. Various deep learning models, such as ARD-Unet, have shown promise in detecting cracks and other defects at the pixel level, thereby increasing the accuracy of infrastructure assessments [14]. YOLO has been utilized in many ways in the road construction industry, mostly for the detection of defects in roads. For instance, YOLO models have been used to detect pavement distress [15][16][17]. However, despite these applications, no YOLO model has yet been trained for the identification of road layers from images, which is another novel approach this paper introduces.

The YOLO framework, introduced by Redmon et al. in 2015 [18], has gained recognition in computer vision for its high accuracy and smaller model size [19]. As its name suggests, YOLO analyzes the entire image in a single pass to extract features. It employs a single CNN on the whole image to simultaneously predict both the bounding boxes and class probabilities [20]. This approach makes the detection process more efficient compared to earlier, more complex methods. The YOLO framework has evolved through several iterations, with each model improving in speed and accuracy. Many developers use YOLO models because they can be trained on a single GPU. YOLO's architecture allows it to achieve high speed and efficiency, making it suitable for real-time processing applications, such as video surveillance [21]. YOLOv11, the latest YOLO model, significantly advances object detection technology by building on its predecessors and introducing innovative enhancements [22]. In benchmark tests, YOLOv11 showed significant improvements in both inference speed and accuracy compared to earlier versions from YOLOv5 through YOLOv10. As illustrated in Figure 1, YOLOv11 not only surpassed these models in achieving higher mean Average Precision (mAP) on the Common Objects in Context (COCO) dataset but also processed data more quickly [23]. The YOLOv11 models demonstrate improved accuracy while minimizing complexity. It is effective in a wide range of computer vision applications and offers faster processing speeds. Importantly, it maintains a balance between accuracy and computational efficiency, along with advanced feature extraction capabilities [23]. YOLOv11 provides five models, including YOLOv11 for Detection, YOLOv11-seg for instance segmentation, YOLOv11-pose for Pose/Keypoints, YOLOv11-obb for Oriented Detection, and YOLOv11-cls for classification [23]. This study suggests a novel approach to using the YOLOv11-seg model for road layer identification and segmentation using images from construction sites.

This study introduces two novel approaches to improving road construction monitoring. This paper presents the innovative application of the YOLOv11-seg model for identifying and segmenting road layers using images, marking its first use in this context. Secondly, while UAVs have previously been used for volume estimations of earthworks and stockpiles, this research extends their application to calculate the volume of road layers. By utilizing four different software tools, this study also identifies the most accurate method for volume calculation. This shows a comprehensive, dual-focused approach that enhances the management and monitoring of road construction projects.

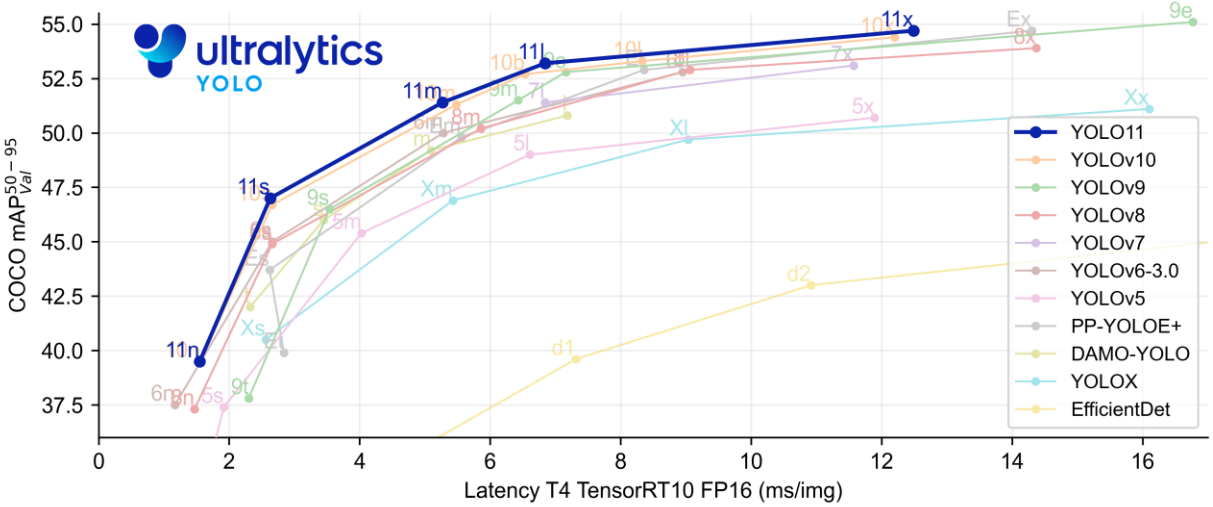


Figure 1: Comparison of YOLOv11 model with earlier YOLO versions in terms of mAP and inference speed on the COCO dataset.

2. YOLOv11 Architecture

The YOLOv11 architecture consists of three important components: Backbone, Neck, and Head.

2.1 Backbone

The backbone is responsible for extracting features from the input images and is an important component of the YOLO architecture. It consists of Convolution layers, C3k2 Block, Spatial Pyramid Pooling Fast (SPPF), and Cross Stage Partial with Spatial Attention (C2PSA) Blocks [24]. The YOLOv11 backbone structure is shown in Figure 2.

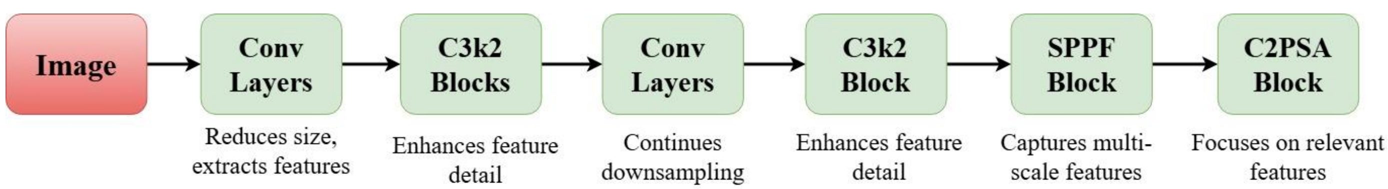


Figure 2: Illustration of YOLOv11 backbone structure highlighting key components such as Convolution layers, C3k2 Block, and SPPF.

2.1.1 Convolution Layers and C3k2 Block

The Convolution layers extract basic features from the input image using various filter sizes. In YOLOv11, these layers reduce the image's spatial dimensions while increasing the number of channels, improving feature detection. A significant update in YOLOv11 is the introduction of the C3k2 block, which replaces the older C2f block found in prior versions [24]. This C3k2 block enhances efficiency by using two smaller convolution layers instead of one large layer, as seen in YOLOv8 [25]. The 'k2' signifies a smaller filter size, enabling faster processing while keeping performance high.

2.1.2 SPPF and C2PSA Blocks

YOLO11 continues to use the SPPF feature from its predecessors while adding the C2PSA block [24] [22]. The SPPF block captures features at multiple scales, while the C2PSA block significantly enhances the model's focus on key areas of the image by using a spatial attention mechanism. This mechanism combines features within the image, enabling YOLO11 to more accurately identify objects.

2.2 Neck

In YOLOv11, the neck serves as a link between the backbone and the head. It plays a key role in integrating and processing features, refining what the backbone extracts before these features are used for predictions. The structure of the neck includes:

2.2.1 C3k2 Block

YOLOv11 introduces the C3k2 block, which replaces the older C2f block used in previous versions. As discussed earlier, the C3k2 block is faster and more efficient, improving the overall performance of the feature aggregation process.

2.2.2 Attention Mechanism

The YOLOv11 uses the C2PSA block to improve spatial attention. This helps the model focus on important parts of an image for better object detection. This feature is not available in YOLOv8, which gives YOLOv11 a potential advantage in accurately detecting smaller or obscured objects [24].

2.3 Head

In YOLOv11, the head is the part of the model that makes its final predictions. The model classifies objects and determines their positions using bounding boxes in the head. The YOLOv11 head structure is shown in Figure 3 and includes:

2.3.1 Upsampling and Concatenation

In YOLOv11, the head uses upsampling and concatenation to improve its ability to detect objects. Upsampling enlarges feature maps so the model can spot smaller objects more effectively. These detailed maps are then combined with features from earlier layers through a process called concatenation. This helps the model represent features more effectively, allowing it to make more accurate predictions.

2.3.2 Feature Processing

In YOLOv11, feature processing plays an important role after upsampling and concatenation. This step uses C3k2 blocks to fine-tune the enhanced features. These blocks bring together information from different layers, which improves the model's ability to identify subtle and complex patterns within images. This process improves the model's accuracy and reliability, enabling it to recognize and classify objects effectively across various settings and backgrounds.

2.3.3 Detection Layer

After the feature processing stage in YOLOv11, the detection layers are engaged. The refined feature maps from various scales are combined to identify objects. The detection layers use these features to accurately predict the locations and classifications of objects within an image (Segment).

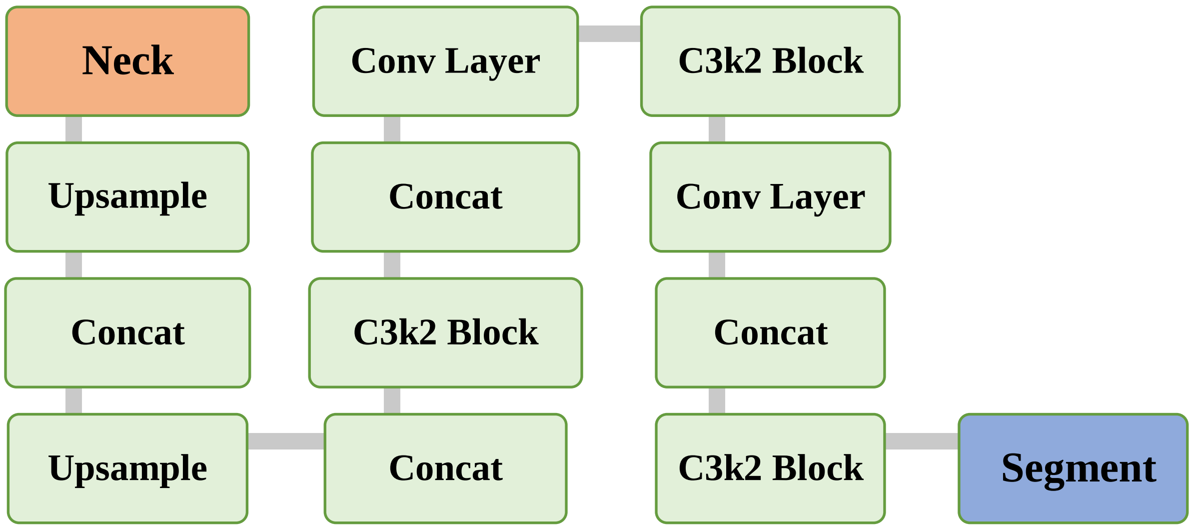


Figure 3: Detailed representation of YOLOv11 head structure showcasing the process of feature integration, object classification, and detection.

3. Volume Calculation Methods: Principles and Applications

The working principle and applications of the four software programs are discussed below:

3.1 Agisoft Metashape

Agisoft Metashape is a photogrammetric software that uses Structure from Motion (SfM) techniques for 3D reconstruction and modeling. By processing overlapping images taken from various angles, it generates dense point clouds and orthomosaics, resulting in precise spatial representations. Metashape's applications range from land cover classification and vegetation monitoring to providing accurate spatial data for agricultural and remote sensing purposes [26]. It's also used to create high-resolution digital surface models and orthomosaics, offering crucial spatial information for construction and architectural projects. The software excels in 3D building and facade modeling, improving visual representation and spatial accuracy assessments. These features make it invaluable for construction monitoring and cultural heritage preservation [27][28]. When integrated with UAV technology, Metashape's capabilities in construction are further enhanced. It processes high-resolution aerial imagery from drones to create detailed 3D models and orthophotos, providing a faster and more accurate alternative to traditional surveying methods [29][30].

3.2 CloudCompare

CloudCompare is an open-source software designed to work with 3D point clouds and meshes. It offers advanced tools for processing, visualizing, and comparing datasets generated through techniques like terrestrial laser scanning and photogrammetry. CloudCompare is widely used in fields such as structural monitoring, archaeological analysis, and environmental modeling. [31][32][33]. In construction, it serves as a powerful tool for volume calculations. By using advanced algorithms, like Multiscale Model to Model Cloud Comparison (M3C2), it enables precise distance measurements between point clouds [34][35], allowing for accurate assessments of material volumes before and after excavation.

3.3 Civil3D

Civil 3D, developed by Autodesk, is a comprehensive civil engineering software that supports a variety of applications, from transportation planning and land development to environmental projects. It's particularly valuable for volume calculations, using advanced algorithms to determine cut and fill volumes from surface models, which are crucial for estimating material requirements in construction projects [36][37]. A key element in Civil 3D is the creation of Triangulated Irregular Network (TIN) surfaces. This process connects data points to form a network of triangles, providing highly accurate representations of existing and proposed terrains, which are important for precise volume calculations. Additionally, integrating Civil 3D with UAV photogrammetry further improves its capabilities [37], allowing for accurate volume estimations for stockpiles and excavation sites.

3.4 WebODM

WebODM is an open-source photogrammetry software with a user-friendly interface, built on the OpenDroneMap (ODM) framework [38][39] that turns aerial images captured by drones into detailed 3D models, including orthophotos, point clouds, and digital elevation models. It works by using a technique called structure-from-motion, which takes overlapping 2D images and converts them into 3D models through steps like dense point cloud generation and mesh creation. In construction, WebODM excels in volume calculations, such as estimating earthwork volumes, stockpile measurements, and excavation assessments. It creates accurate 3D models and point clouds from aerial images, making it especially useful for large-scale projects like highways and ports, where traditional methods can struggle with time and accuracy.

4. Methodology

4.1 Detection Model

The process of training the detection model included data collection, processing, and model training.

4.1.1 Data Collection and Processing

Using a DJI Phantom 4 Pro drone, we captured high-quality, overlapping photographs of various road layers at multiple construction sites, including the Military College of Engineering in Risalpur and DHA Quetta. Additionally, to enhance the diversity and robustness of our dataset, we gathered relevant images from the internet. The collection consisted of approximately 1,000 images, which included aerial, front, and side views. Importantly, the images from the Military College of Engineering were georeferenced, and not only did they serve as a critical component for training our detection model, but they were also used in the volume calculation phase of our study. To give a clearer view of the sites and the types of images collected, Figure 4a - 4c showcases a selection of the photographs captured from these locations. After collection, the RGB images underwent two crucial processing steps to prepare them for training the model. First, we annotated more than

1,000 images using the Labelme tool, ensuring precise labeling of road layers. Next, we converted the annotation data from JSON format, which is incompatible with YOLO, into a format acceptable for YOLO, see Figure 5, using the labelme2yolo tool. This process ensured that the data could be effectively used for training.

(a)

(b)

(c)

(d)

Figure 4: Samples of RGB images captured from different angles, illustrating data collection diversity.

Class x_center y_center width height

Figure 5: YOLO accepted format

4.1.2 Model Training

The YOLOv11-seg model was trained using annotated images on Google Colab, which provided access to a powerful GPU to handle the processing demands efficiently. We trained the model for 350 epochs using an image size of 640 for both training and validation, with a batch size of 8, which was used for less computational power requirement. The dataset was divided into 70% for training and 30% for validation to assess the model's performance accurately. To increase the dataset's diversity and enhance the model's generalization capabilities, we applied various data augmentation techniques using albumentations. These included slight blurring, median blurring, converting images to grayscale, and applying Contrast Limited Adaptive Histogram Equalization (CLAHE) to adjust image contrast. Each augmentation was applied with a probability of 0.01, ensuring subtle yet effective variations. The optimizer used was AdamW, with a momentum of 0.9, a weight decay factor of 0.0005 to prevent overfitting and a learning rate of 0.001429. We used eight data loader workers to effectively manage data throughput during training. Additionally, early stopping criteria were set, and regular checkpoints were saved to avoid overfitting, where the model would cease the training if the validation stopped improving. Table 1 summarizes all the hyperparameters which were used in the training process.

Table 1: Summary of key hyperparameters used during YOLOv11 model training, including epochs, learning rate, and image size.

Hyperparameter	**Value**
Epochs	350
Image Size	640
Batch Size	8 images
Learning rate	0.001429
Momentum	0.9
Weight decay	0.0005

4.2 Volume Calculation

The methodology for calculating volume comprised three key stages: data collection, dense point cloud generation, and volume computation using four different software tools.

4.2.1 Data Collection

The RGB images of the MCE site captured by the drone for the YOLO model were also used for volume calculation. These images were intended for dual use: to train the YOLO model and to facilitate volume calculation. To ensure high accuracy in the volume measurements, ground control points (GCPs) were established and visibly marked within the images. For example, two GCPs visible in the Figure were recorded with their own coordinates: one at Latitude 34.049655° N, Longitude 71.989204° E, and Altitude 212.229064 meters; the other at Latitude 34.049569° N, Longitude 71.989422° E, and Altitude 211.718859 meters. This illustrates that each GCP has its specific location data. This step was essential because GPS data can have errors, which could lead to inaccuracies. GCPs provided more reliable reference points, significantly improving the precision of our volume calculations.

Figure 6: Georeferenced RGB image with GCPs established for improved accuracy in volume calculations.

4.2.2 Dense Point Cloud Generation

Using Agisoft Metashape, we generated point clouds from the RGB images through photogrammetric reconstruction. This technique allowed us to create detailed 3D models of the road surface, capturing it both before the new layer was laid and after. GCPs were crucial for accurately aligning the point cloud with real-world coordinates, ensuring precise georeferencing. This accuracy was essential for creating a reliable 3D model. After georeferencing, we post-processed the point clouds to generate dense point clouds. This step enhanced the level of detail, providing a thorough representation of the road surface. For example, Figure 7 shows the dense point cloud of a specific site before any layer application, with a red rectangle indicating the area designated for layer application. Figure 8 shows the same location after the application of aggregate base and subgrade layers. Comparing these examples helps us evaluate the changes in the road surface resulting from the new layer application.

Figure 7: Dense point cloud of the construction site prior to the application of road layers, illustrating initial surface conditions.

Figure 8: Dense point cloud post-application of aggregate base and subgrade layers, showcasing layer addition.

4.2.3 Volume Computation

The dense point clouds generated by Agisoft Metashape were used to calculate the volume of the road layers across multiple software tools, including Metashape itself, CloudCompare, and Civil 3D, all of which relied on the Metashape-generated point cloud for volume computation. However, it's important to note that Metashape used the dense point cloud of the site after the layer application, focusing on the post-application surface. In contrast, Civil 3D and CloudCompare used both pre- and post-application dense point clouds, allowing for a comparative analysis to determine the volume. WebODM, on the other hand, generated its own post-application dense point cloud to determine the volume. The following sections detail the volume calculation process for each tool

4.2.3.1 Agisoft Metashape:

In Agisoft Metashape, polygons were drawn around the boundaries of each road layer, the subgrade layer, the aggregate base, and the asphalt (see Figure 9a—Figure 9c). These boundaries were created to obtain the correct volume of each layer. After the layer boundaries were set, the software calculated the volume of each layer with regard to elevation variation within these polygons. This way, more accurate estimations of the volume of each layer were achieved with reference to the detailed 3D models obtained from the dense point clouds.

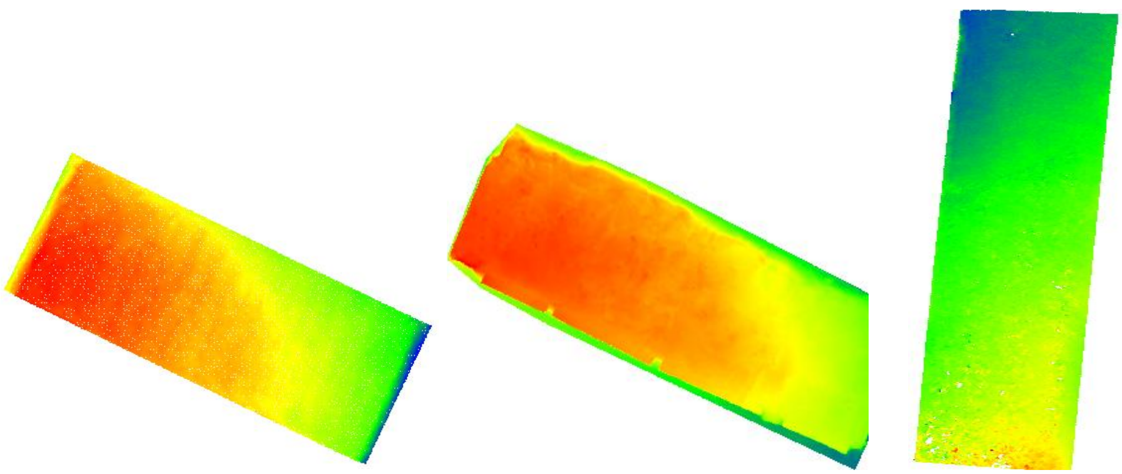
Figure 9: Polygon Boundaries around Aggregate Base, Subgrade and Asphalt

(a) (b) (c)

4.2.3.2 CloudCompare:

In CloudCompare, we determined the volume of the road layer by comparing the point cloud of the construction site before the layer application with the point cloud of the site after the layer application. We isolated the area of interest by removing irrelevant environmental elements. Utilizing CloudCompare's volumetric analysis tools, we calculated the volume of material in the layer by using a grid-based approach with a step size of 0.009. Figures 10A to 10C illustrate the elevation differences across the layers, with blue indicating the least variation, red indicating the largest variation, and green representing the intermediate values.

Figure 10: Elevation differences captured in CloudCompare across (a) Aggregate Base, (b) Subgrade, and (c) Asphalt, highlighted using a color gradient from blue (least variation) to red (largest variation).



(a) (b) (c)

4.2.3.3 Civil 3D:

In Civil 3D, we worked with point clouds converted into RCP format via Autodesk Recap from the original Agisoft Metashape outputs. These dense point clouds, representing the pre- and post-construction phases, were used to create TIN surfaces for both stages. By overlaying these surfaces, we could analyze the changes in each layer, including subgrade, aggregate base, and asphalt. This comparison allowed us to determine the volume of material added or removed during construction. Figure 11a – 11c shows the overlapped surfaces of the layers of both stages.

(a) (b) (c)

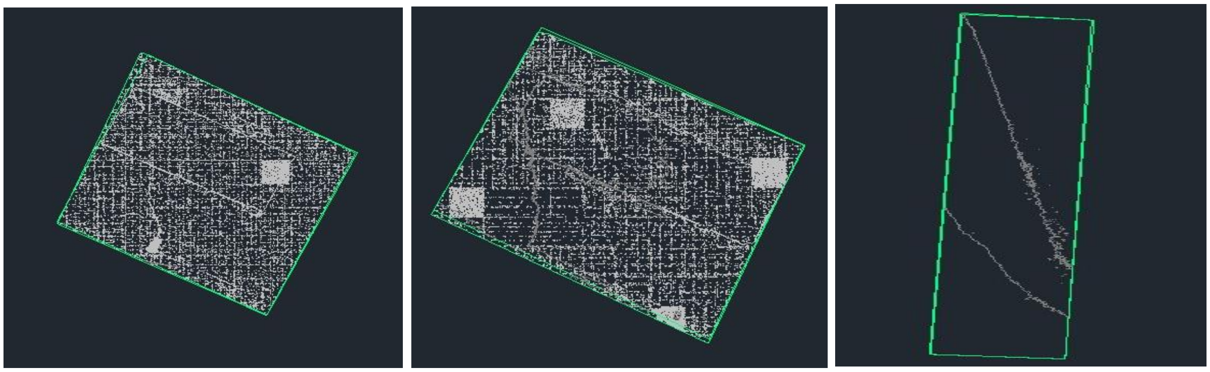


Figure 11: Overlapped Surface of (a) Aggregate Base, (b) Subgrade and (c) Asphalt before and after construction

4.2.3.4 WebODM:

In WebODM, we processed RGB drone images taken after the application of the road layer to create a detailed point cloud and 3D model of the construction site. This post-application point cloud was generated with high geospatial accuracy through the integration of GCPs. WebODM allowed us to define the boundaries of the road layers within this point cloud, facilitating precise volume calculations based on elevation data. This approach provided us with a comprehensive understanding of the material volumes involved in the construction process, enhancing the accuracy of our volumetric analysis.

5. Results and Discussion

5.1 Metrics Used

To assess the performance of YOLOv11, we used several key metrics to help us understand the model's accuracy and efficiency. The primary metrics included:

5.1.1 Precision and Recall

Precision measures the accuracy of the model in identifying only relevant objects and is calculated as the ratio of true positive detections to the total number of detections by the model.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positives}} \tag{1}$$

Recall measures the accuracy of the model in identifying all relevant objects within the dataset and is calculated as the ratio of true positive detections to the total number of actual positives in the dataset.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive+False Negatives}} \tag{2}$$

5.1.2 F1-Score

F1-Score combines the harmonic mean of precision and recall into a single value, providing a balanced measure of the model's accuracy.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

5.1.3 Mean Average Precision (mAP)

Mean Average Precision (mAP) is a widely used metric for evaluating object detection models. It calculates average precision at different Intersection over Union (IoU) thresholds. There are two primary mAP calculations reported:

- **mAP@0.5**: This metric measures average precision at an IoU threshold of 0.5, primarily focusing on the model's accuracy in classifying objects.
- **mAP@0.5:0.95**: This average precision is calculated across various IoU thresholds, ranging from 0.5 to 0.95, ensuring a thorough assessment of the model's localization and classification abilities.

$$\text{mAP} = \frac{1}{\text{Number of Classes}} \times \sum_{i=1}^{N} \text{Average Precision } for\ class\ i\ (AP_i) \quad (4)$$

5.1.4 Confusion Matrix

The confusion matrix illustrates the model's performance in predicting various road layer classes by comparing predicted labels with actual labels. We used a normalized confusion matrix to clearly indicate the model's strong performance areas and its frequent mistakes.

5.2 YOLOv11 Results

5.2.1 Confusion Matrix Evaluation

The normalized confusion matrix, depicted in Figure 12, provides a detailed view of YOLOv11's classification accuracy for different road layer types. Key insights from the analysis include:

- **Aggregate Base**: Correctly predicted 95% of the time (0.95). However, there is some confusion with Asphalt.
- **Subgrade**: Correctly predicted 83% of the time (0.83). However, there is some confusion with the aggregate base.
- **Asphalt**: Correctly predicted 95% of the time (0.95).

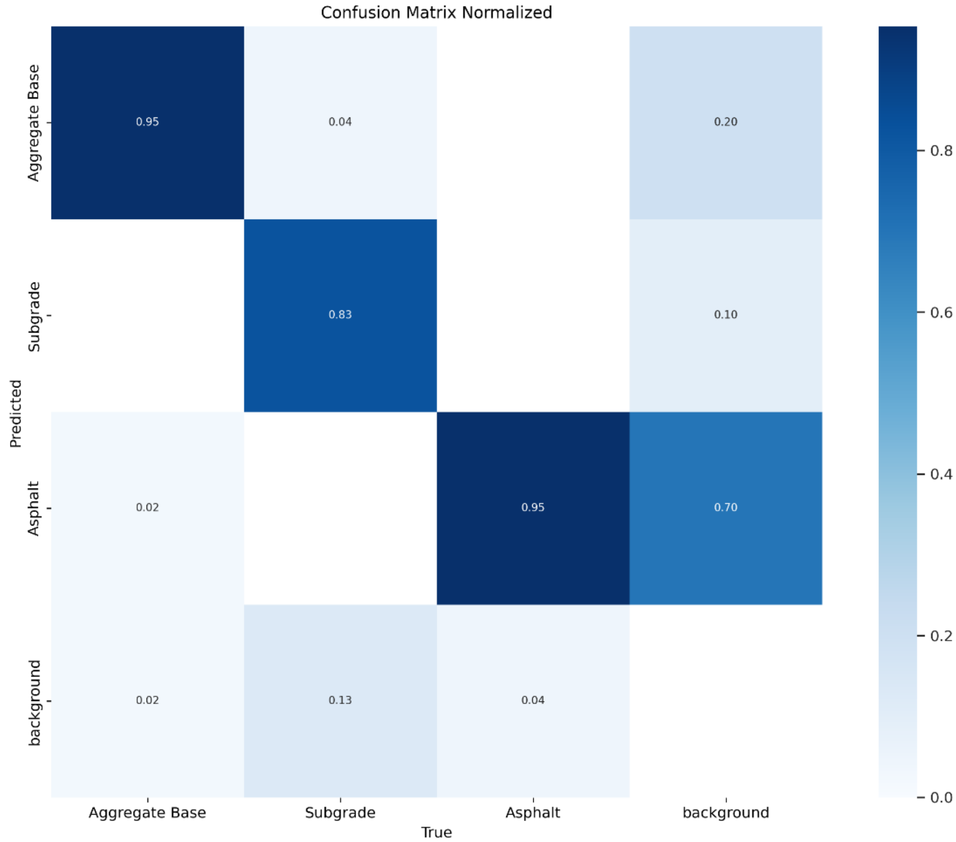


Figure 12: Normalized Confusion Matrix.

5.2.2 F1 Confidence Curve

The model achieves its highest F1 score of 0.92 at a confidence threshold of around 0.779, see Figure 13, indicating an optimal balance of precision and recall at this level. This threshold is well-suited for scenarios where a high F1 score is crucial, as it provides an effective balance between precision and recall across all classes.

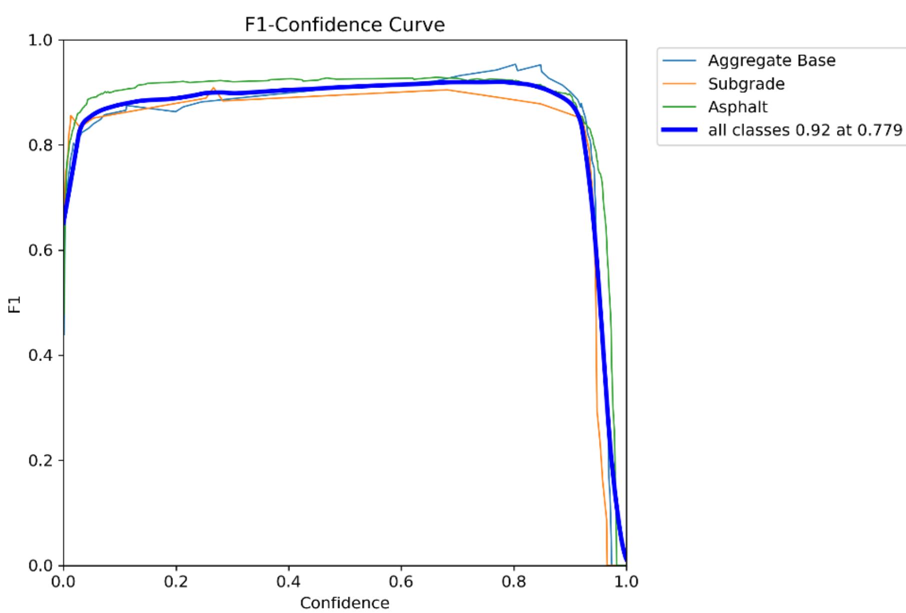


Figure 13: F1 Confidence Curve

5.2.3 Precision Confidence Curve

The precision-confidence curve shows how precision varies at different confidence levels for each class. The model consistently achieves high precision, peaking at 1.0 when the confidence threshold is set to 1.0, see Figure 14. The "Subgrade" class demonstrates particularly high precision, whereas "Aggregate Base" and "Asphalt" have slightly lower precision, especially at lower confidence levels. This suggests that the model is most reliable for "Subgrade," with precision improving for all classes as the confidence threshold increases.

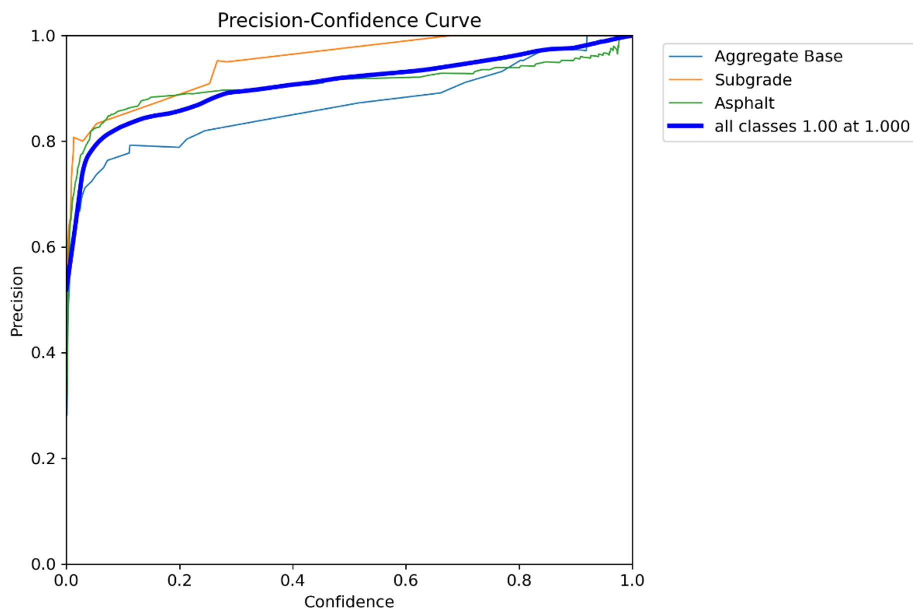


Figure 14: Precision-Confidence Curve.

5.2.4 Precision Recall Curve

The precision-recall curve illustrates the model's capability to balance precision and recall across different classes. The model achieves high precision and recall for both "Aggregate Base" and "Asphalt". However, for "Subgrade," precision decreases as recall increases, indicating difficulty in maintaining precision at higher recall levels. The overall mAP@0.5 IoU reflects a well-maintained balance of precision and recall across the dataset. Figure 15 offers a comprehensive overview of each class's performance.

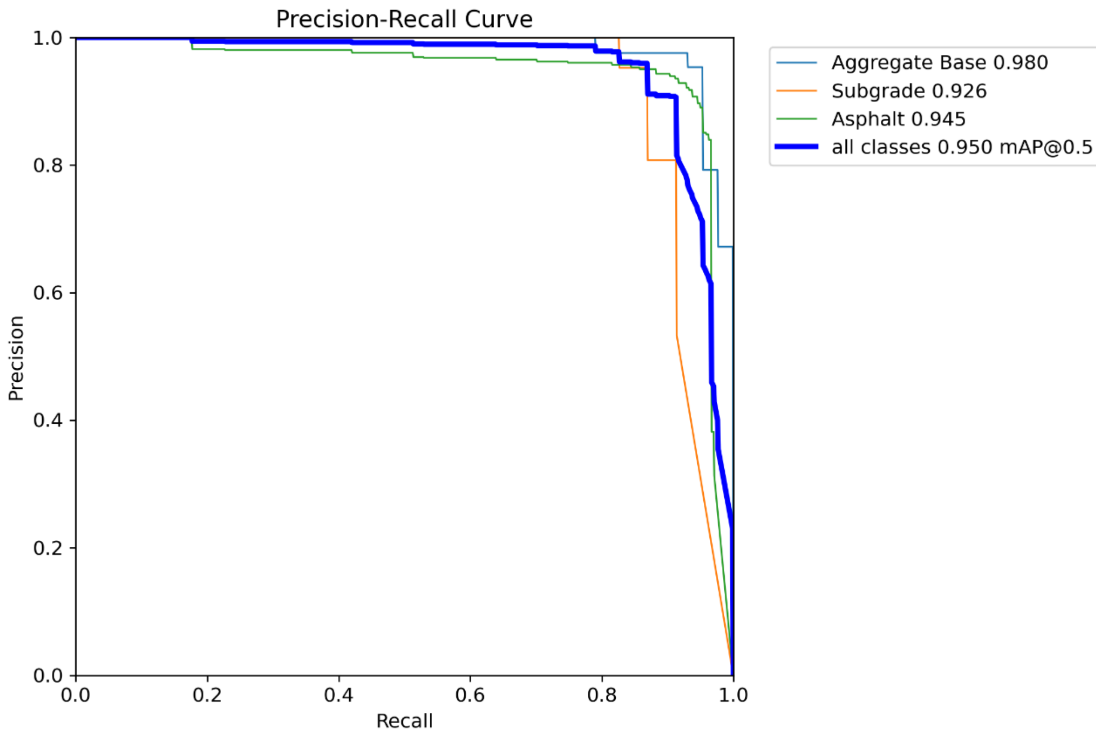


Figure 15: Precision-Recall Curve

5.2.5 Recall Confidence Curve

The recall-confidence curve shows how recall varies at different confidence levels. The model attains a high recall of 0.96 with a low confidence threshold, as depicted in Figure 16. This shows that the model effectively identifies most instances across all classes when the threshold is set lower. As the confidence threshold increases, recall decreases, showing the balance between achieving high precision and maintaining a high recall rate.

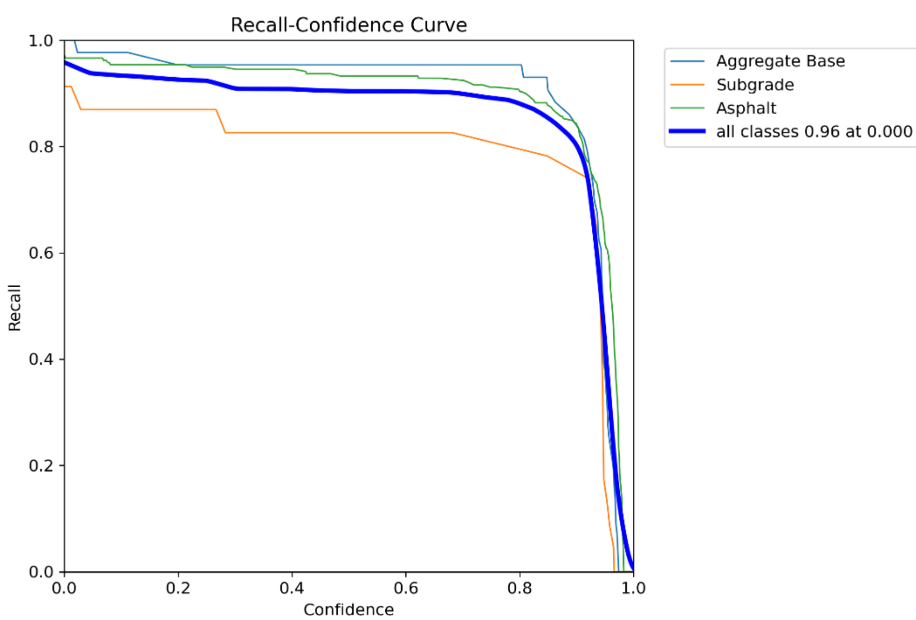


Figure 16: Recall-Confidence Curve

5.2.6 Training and Validation Loss

The training and validation loss curves in Figure 17 show how YOLOv11 learned during training. Both losses went down steadily, showing that the model learned well without overfitting. YOLO11's consistent performance in terms of precision and recall proves its reliability in road layer detection tasks.

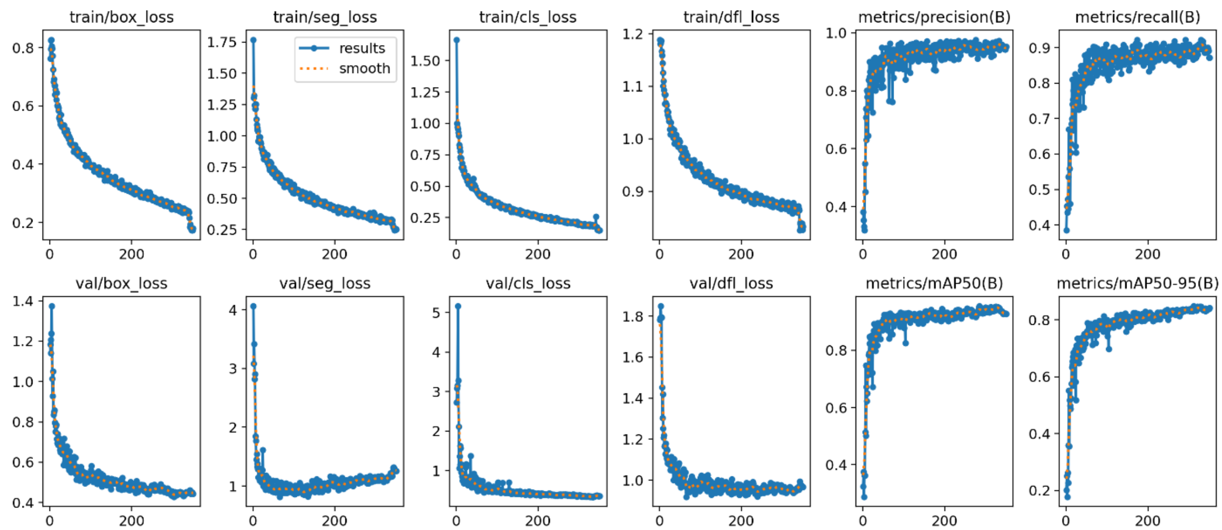


Figure 17: Training and validation loss curves indicating YOLOv11's learning progression and model convergence during training.

5.2.7 Model Detection Results

The trained YOLOV11 model was used to detect road layers in images not included in the training set, ensuring that the model's performance was evaluated on entirely new data. YOLOv11 processes an image by dividing it into a grid and then predicting bounding boxes and class probabilities for each grid cell in a single pass. This allows the model to efficiently identify multiple objects and their classes within an image. Figures 18a and 18b presents the detection results from the trained YOLOv11-seg model applied to test images. Figures 19a and 19b display the detection results using the YOLOv11 model on images that were not part of the training set. Meanwhile, Figures 20a and 20b illustrate the detection results for "Asphalt" using the same model on images that were not part of the training set.

(a) (b)

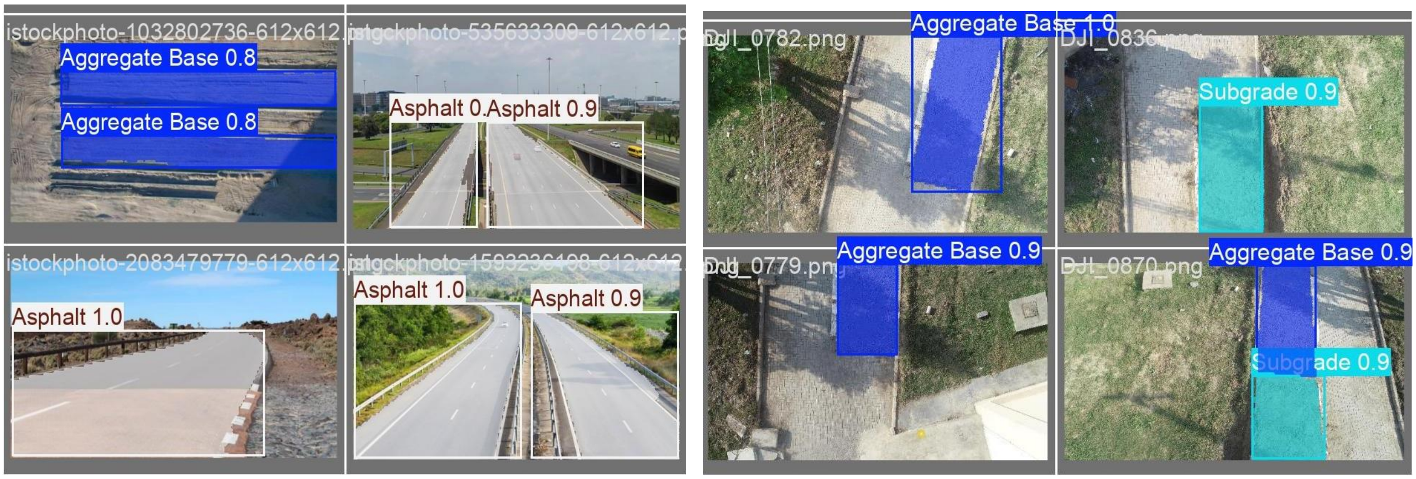


Figure 18a – 18b: Detection results on test images using YOLOv11.

(a)

(b)

Figure 19a – 19b: Detection results using the YOLOv11 model on images not included in the training.

(a) (b)

Figure 20a – 20b: Detection results of asphalt layer using the YOLOv11 model on images not included in the training.

5.3 Volume Calculation

To assess the accuracy and reliability of these software tools, the calculated subgrade volumes from Agisoft Metashape, CloudCompare, Civil 3D, and Web ODM were compared with the actual volume laid on the site (see Table 2).

Table 2: Comparison of calculated and actual road layer volumes using different software tools, highlighting accuracy levels for subgrade, aggregate base, and asphalt layers.

Road Layer	Software	Calculated Volume (m^3)	Actual Volume (m^3)	%Accurate
Subgrade	**Agisoft Metashape**	4.731	4.59	96.93
	CloudCompare	4.636		98.99
	Civil3D	4.75		96.51
	WebODM	4.597		99.85
Aggregate Base	**Agisoft Metashape**	2.426	2.26	92.65
	CloudCompare	2.395		94.03
	Civil3D	2.63		83.62
	WebODM	2.262		99.9
Asphalt	**Agisoft Metashape**	8.083	8.17	98.93
	CloudCompare	8.916		90.86
	Civil3D	9.37		85.31
	WebODM	8.182		99.85

The comparison of calculated and actual volumes using different software tools highlights the varying degrees of accuracy achieved by each method. The evaluations indicated that WebODM was the most accurate, and Agisoft Metashape and CloudCompare were second most accurate compared to Civil 3D. These results also indicate that WebODM outperforms other methods in terms of volume calculations, suggesting its potential as a preferred tool for precise and reliable calculations in road construction projects using UAVs. Following this, we present a detailed comparison of the advantages and disadvantages observed in our study for each software tool in Table 3. This table is based on our practical experience and observations during the course of our research.

Table 3: Advantages and disadvantages of volume calculation tools, including Agisoft Metashape, CloudCompare, Civil 3D, and WebODM.

Software	Advantages	Disadvantages
Agisoft Metashape	- High precision in creating dense point clouds and 3D models from UAV imagery. - User-friendly interface for defining boundaries for volume calculations. - Integration with GCPs improves accuracy. - Effective for elevation-based volume calculations and photogrammetry.	- Computationally intensive; requires significant processing power and time for large datasets.
CloudCompare	- Open-source and cost-free, making it accessible for budget-sensitive projects. - Specialized in point cloud analysis with advanced volume calculation algorithms (e.g., grid-based and concave hull methods). - Effective for comparing pre- and post-construction data.	- Complex interface, which can be challenging for new users. - Requires manual setup (e.g., cropping areas of interest), increasing the chance of user error.
Civil 3D	- Tailored for civil engineering with robust tools for grading, surface modeling, and cross-sections. - Excels in cut-and-fill volume calculations and aligns with CAD workflows. - Integrates seamlessly with BIM for large-scale infrastructure projects.	- High licensing cost and steep learning curve for new users. - Limited photogrammetry capabilities; cannot process dense point clouds without preprocessing.
WebODM	- Open-source, user-friendly, and suitable for budget-conscious projects. - Capable of processing UAV imagery for generating dense point clouds and accurate volume calculations. - Flexible deployment options (local or cloud-based processing).	- Limited engineering-specific tools for advanced analyses like cut-and-fill. - Struggles with very large datasets on low-spec systems. - Lacks some customization features and integration capabilities available in more specialized software like Civil 3D or Metashape.

6. Conclusion

This study introduces a novel method for detecting road layers and calculating their volume by combining drone technology with artificial intelligence, thereby improving the monitoring of road construction projects. We utilized high-resolution images captured by a DJI Phantom 4 Pro drone along with images taken by the internet to train the YOLOv11-seg model. This model was trained to classify and segment different road layers, including subgrade, asphalt, and aggregate base, directly from the images. The model demonstrated impressive accuracy in identifying road layers, achieving 95% accuracy for both the aggregate base and asphalt layers and 83% accuracy for the subgrade layer. It also performed exceptionally well on other important metrics, reaching perfect precision at the highest confidence level. Additionally, it achieved an F1 score of 0.92 at a confidence threshold of 0.779 and a recall of 0.96 at lower confidence levels. This consistent performance highlights the model's capability to accurately and reliably identify road layers, ensuring effective and comprehensive monitoring of road construction projects.

The high-resolution images captured by the UAV not only aided in training the model but also enabled accurate volume calculations. To improve precision, Ground Control Points (GCPs) were established on-site since GPS coordinates alone do not offer sufficient accuracy. In this study, we presented a new method for calculating the volume of road layers using UAV technology. Additionally, we compared the results from four different software tools to determine the most effective approach. Among these, WebODM proved to be the most accurate, closely matching the actual volumes on construction sites with an accuracy of 99.85% for subgrade and asphalt, and 99.9% for aggregate base. Agisoft Metashape closely followed, showing accuracies of 96.93% for subgrade, 92.65% for aggregate base, and 98.93% for asphalt. Other tools also performed well; CloudCompare achieved a 98.99% accuracy for subgrade, 90.86% for asphalt and 94.03% for aggregate base, while Civil3D showed accuracies of 96.51% for subgrade, 83.62% for aggregate base, and 85.31% for asphalt. This approach minimizes human error, enhances data security, and ensures consistent construction progress monitoring. As a result, adopting these technologies can lead to improved project management, cost savings, and better road construction quality.

This research raises several important questions and provides opportunities for further investigation. One key area for development is improving the current YOLO model to detect and analyze more complex aspects of road construction. Although this paper has focused on three primary layers, including subgrade, aggregate base course, and asphalt course, road construction, in reality, often involves a more complex layering structure. Enhancing the model's ability to identify and segment these additional layers could transform it into a more effective monitoring tool. Furthermore, while this study used high-resolution images, integrating LiDAR technology could further improve the volume calculation process. LiDAR technology offers high accuracy, which can lead to more precise volume measurements. By applying the findings from this research to real-world projects, we could significantly reduce the need for frequent site visits and the number of surveyors required. This would optimize both the cost and efficiency of monitoring road construction.

Data Availability

Data was generated during the study and is available at https://figshare.com/s/8c22df61ae4f3bfaafd5

Abbreviations

YOLO – You Only Look Once

UAV – Unmanned Aerial Vehicle

GNSS - Global Navigation Satellite System

AI – Artificial Intelligence

LiDAR - Light Detection and Ranging

CNN - Convolutional Neural Network

MRF - Markov Random Fields

mAP - mean Average Precision

COCO - Common Objects in Context

SPPF - Spatial Pyramid Pooling Fast

C2PSA - Cross Stage Partial with Spatial Attention

SfM - Structure from Motion

M3C2 - Multiscale Model to Model Cloud Comparison

TIN - Triangulated Irregular Network

ODM – OpenDroneMap

CLAHE - Contrast Limited Adaptive Histogram Equalization

GCP - Ground Control Points

IoU - Intersection over Union

Acknowledgement

We would like to extend our gratitude to the Military College of Engineering (NUST, Risalpur, Pakistan) for their support and resources, which made this research possible. Special thanks to the technical staff and surveyors, who provided invaluable assistance during the data collection processes.

Declarations

Competing Interests: The authors declare no competing interests.

Funding: The authors did not receive support from any organization for the submitted work.

References

[1] Vick S and Brilakis I 2023 “A Review of Linear Transportation Construction Progress Monitoring Techniques” *Apollo - University of Cambridge Repository*

[2] Srinavin K, Rattanapongwanich S, Kusonkhum W, Leungbootnak N and Charnwasununth P 2020 “Construction progress evaluation using 3-D model based on point cloud” *International Journal of Structural and Civil Engineering Research* **9(4)** 278

[3] Tjandra K D, Irawan F G, Nugraha P and Sunindijo R Y 2022 “Drone Readiness in the Indonesian Construction Industry” *Construction Economics and Building* **22(4)** https://doi.org/10.5130/AJCEB.v22i4.8161

[4] Small E P, Hendricks J and Woodacre K 2019 “Effectiveness of Drone-Based Photogrammetry for On-Site Quantity Assessment” *Creative Construction Conference 2019*, 292 https://doi.org/10.3311/CC C2019-042

[5] Ajayi O.G. and Ajulo J. 2021 “Investigating the Applicability of Unmanned Aerial Vehicles (UAV) Photogrammetry for the Estimation of the Volume of Stockpiles” *Quaestiones Geographicae* **40(1)** 25-38 https://doi.org/10.2478/quageo-2021-0002

[6] Fang L, You Z, Shen G, Chen Y and Li J 2022 “A joint deep learning network of point clouds and multiple views for roadside object classification from lidar point clouds” *ISPRS Journal of Photogrammetry and Remote Sensing* **193** 115-136

[7] Lee S, Han D and Song M 2022 “Calculation and comparison of earthwork volume using unmanned aerial vehicle photogrammetry and traditional surveying method” *Sensors and Materials* **34(12)** 4737

[8] Lee K and Lee W H 2022 “Earthwork volume calculation, 3D model generation, and comparative evaluation using vertical and high-oblique images acquired by unmanned aerial vehicles” *Aerospace* **9(10)** 606

[9] Kim J, Lee S, Seo J, Lee D-E, and Choi HS 2021 “The Integration of Earthwork Design Review and Planning Using UAV-Based Point Cloud and BIM” *Applied Sciences* **11(8)** 3435 https://doi.org/10.3390/app11083435

[10] Mora R, Martín-Jiménez J A, Lagüela S and González-Aguilera D 2021 “Automatic point-cloud registration for quality control in building works” *Applied Sciences* **11(4)** 1465

[11] Guo X. 2023 “Artificial Intelligence in the Construction Sector” *Applied and Computational Engineering* **28** 1-6 https://doi.org/10.54254/2755-2721/28/20230004

[12] He X, Tang Z, Deng Y, Zhou G, Wang Y and Li L 2023 “UAV-based road crack object-detection algorithm” *Automation in Construction* **154** 105014

[13] Lei G, Jiangdong S, Zhitao X, Fang Z and Jun W 2018 “Combining CNN and MRF for road detection” *Computers & Electrical Engineering* **70** 895-903

[14]Gao Y, Cao H, Cai W and Zhou 2020 “Pixel-level Road crack detection in UAV remote sensing images based on ARD-Unet” *Reports on Geodesy and Geoinformatics* **110(1)** 21-28

[15] Li J, Yuan C, and Wang X 2023 “Real-time Instance-level Detection of Asphalt Pavement Distress Combining Space-to-Depth (SPD) YOLO and Omni-scale Network (OSNet)” *Automation in Construction* **155** 105062 https://doi.org/10.1016/j.autcon.2023.105062

[16] Ghosh R and Smadi O 2021 “Automated Detection and Classification of Pavement Distresses using 3D Pavement Surface Images and Deep Learning” *Transportation Research Record* **2675(9)** 1359-1374 https://doi.org/10.1177/03611981211007481

[17] Zhang Y, Zuo Z, Xu X, Wu J, Zhu J, Zhang H, Wang J, and Tian Y 2022 “Road Damage Detection Using UAV Images Based on Multi-level Attention Mechanism” *Automation in Construction* **144** 104613 https://doi.org/10.1016/j.autcon.2022.104613

[18] Redmon J, Divvala S, Girshick R and Farhadi A 2016 “You only look once: Unified, real time object detection” IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[19] Terven J, Cordova-Esparza D and Romero-González J-A 2023 “A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS” *Machine Learning and Knowledge Extraction* **5(4)** 1680-1716

[20] Du J 2018 “Understanding of Object Detection Based on CNN Family and YOLO” *Journal of Physics: Conference Series* **1004** 012029 https://doi.org/10.1088/1742-6596/1004/1/012029

[21] Lavanya G and Pande S D 2023 “Enhancing Real-time Object Detection with YOLO Algorithm” *EAI Endorsed Transactions on Internet of Things* **10** https://doi.org/10.4108/eetiot.4541

[22] Khanam R and Hussain M 2024 “YOLOv11: An Overview of the Key Architectural Enhancements” https://doi.org/10.48550/arXiv.2410.17725

[23] Ultralytics 2024 “Ultralytics YOLOv11” Accessed: October 21, 2024, from https://docs.ultralytics.com/models/yolo11/

[24] Ghosh A 2024 “Ultralytics YOLO11: Faster Than You Can Imagine!” Accessed: October 21, 2024, from https://learnopencv.com/yolo11/

[25]Francesco Jacob Solawetz 2023 “What is YOLOv8? A Complete Guide.” Accessed: 06 Nov 2024, from https://blog.roboflow.com/what-is-yolov8/

[26] Sefercik U G, Kavzoglu T, Colkesen I, Adali S, Dinc S, Nazar M and Ozturk M Y 2021 "Land Cover Classification Performance of Multispectral RTK UAVs" *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLVI-4/W5-2021** 489–492 https://doi.org/10.5194/isprs-archives-XLVI-4-W5-2021-489-2021

[27] Rahman A D M and Cahyono A B 2023 "Analysis of 3-D Building Modeling Using Photogrammetric Software: Agisoft Metashape and Micmac" *IOP Conf. Ser.: Earth Environ. Sci.* **1276** 012044 https://doi.org/10.1088/1755-1315/1276/1/012044

[28] Taddia Y, González-García L, Zambello E and Pellegrinelli A 2020 "Quality Assessment of Photogrammetric Models for Façade and Building Reconstruction Using DJI Phantom 4 RTK" *Remote Sensing* **12(19)** 3144 https://doi.org/10.3390/rs12193144

[29] Pinatik N Y and Papilaya F S 2024 "Pengolahan Foto Udara UAV (Unmanned Aerial Vehicle) Menggunakan Software Agisoft Metashape" *Jurnal Perangkat Lunak* **6(1)** 1–11 https://doi.org/10.32520/jupel.v6i1.2838

[30] Kamal A, `Aziz A A, Ahmad M A, Ahmad M J and Srivanit M 2023 "3D Model of Tunku Tun Aminah Library Using Unmanned Aerial Vehicle (UAV)" *International Journal of Sustainable Construction Engineering and Technology* **14(5)** 408–415 https://publisher.uthm.edu.my/ojs/index.php/IJSCET/article/view/15328

[31] Jennings J, Hawkins A, Sellen A and Spence Morrow G 2022 "Fakers and Mold-Makers: The Use of Structured Light Scanning to Detect Forgeries of Pre-Hispanic Effigies from Oaxaca" *Journal of Conservation and Museum Studies* **20(1)** 3 https://doi.org/10.5334/jcms.216

[32] Gura D, Karamysheva E and Pshidatok S 2024 "Using CloudCompare Software Editing Tools for Processing a Three-Dimensional Point Cloud of an Urban Development Site" *ITM Web of Conferences* **59** 02008 https://doi.org/10.1051/itmconf/20245902008

[33] Nagle-McNaughton T and Cox R 2020 "Measuring Change Using Quantitative Differencing of Repeat Structure-From-Motion Photogrammetry: The Effect of Storms on Coastal Boulder Deposits" *Remote Sensing* **12(1)** 42 https://doi.org/10.3390/rs12010042

[34] Esposito G, Mastrorocco G, Salvini R et al. 2017 "Application of UAV Photogrammetry for the Multi-Temporal Estimation of Surface Extent and Volumetric Excavation in the Sa Pigada Bianca Open-Pit Mine, Sardinia, Italy" *Environmental Earth Sciences* **76** 103 https://doi.org/10.1007/s12665-017-6409-z

[35] Panagiotopoulou A, Wallace C A B, Ragia L and Moullou D 2023 "Change Detection between Retrospective and Contemporary 3D Models of the Omega House at the Athenian Agora" *Heritage* **6(2)** 1645–1679 https://doi.org/10.3390/heritage6020088

[36] Mohamed A M 2019 "Volume Calculation of Irregular Object Using Multiple Software Packages" *Journal of Al-Azhar University Engineering Sector* **14(50)** 75–86 https://doi.org/10.21608/auej.2019.28507

[37] Halim A E A, Ab Rahman A A, Abdul Maulud K N, Nekmat M S and Mukhlisin M 2023 "Assessment of Difference in Structure from Motion for Stockpile Volume Estimation Using UAV Approach" *IOP Conf. Ser.: Earth Environ. Sci.* ***1240*** *012002* https://doi.org/10.1088/1755-1315/1240/1/012002

[38] Vacca G 2019 "Overview of Open Source Software for Close Range Photogrammetry" *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLII-4/W14** 239–245 https://doi.org/10.5194/isprs-archives-XLII-4-W14-239-2019

[39] WebODM Authors “WebODM” Accessed: November 10, 2024, from https://github.com/OpenDroneMap/WebODM/