ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING POTENTIAL AS TOOLS FOR GEOREFERENCING AND FEATURES DETECTION USING UAV IMAGERY

Amr M. Elsheshtawy[1], Mohamed A. Amasha[1], Doaa A. Mohamed[2], Mohamed E. Abdelmomen[2], Mahmoud A. Ahmed[1], Mohamed A. Diab[3], Mostafa M. Elnabawy[1], Dina M. Ahmed[4], Ammar R. Aldakiki[1]

1Civil Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt;
2 Computer Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt;
3 Mechanical Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt;
4 Electrical Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt.

KEY WORDS: UAVs, Photogrammetry, Remote sensing, Special resolution, Artificial intelligence, Machine learning

ABSTRACT:

While Satellite imagery holds the advantage of encompassing expansive geographical regions. spanning square kilometers, its Spatial Resolution (SR) might prove inadequate for specific tasks. Conversely, Unmanned Aerial Vehicles (UAVs) excel in capturing high-resolution images with spatial resolutions in the range of a few centimeters or even millimeters. However, the accuracy of sensor locations during UAV flights is non-accurate enough by the Global Navigation Satellite Systems (GNSS) technology onboard. One of the key objectives of this research is to evaluate a technique aimed at generating precise sensor locations. This technique employs raw data from the drone's GNSS receiver and minimum Ground Control Points (GCPs) placed within a 2-meter diameter circle in the study area. The goal is to achieve accurate Digital Elevation Models (DEM) and orthomosaic images.

Another focus of this research is on addressing challenges related to road lane detection. This is achieved through the enhancement of the You Only Look Once (YOLO) v3 algorithm. The proposed approach optimizes grid division, detection scales, and network architecture to enhance accuracy and real-time performance. The experimental results showcase an impressive 92.03% accuracy with a processing speed of 48 frames per second (fps), surpassing the performance of the original YOLOv3. In the rapidly evolving landscape of Artificial Intelligence (AI) and drone technology, this investigation underscores both the potential and complexities inherent in utilizing advanced AI models, such as YOLOv8, for building detection using UAV and satellite imagery. Furthermore, the research delves into robustness and real-time capabilities within building detection algorithms. The outlined strategy encompasses precise pre-processing, Field-Programmable Gate Array (FPGA) validation, and algorithm refinement. This comprehensive framework aims to elevate feature detection in intricate scenarios, ensuring accuracy, real-time efficiency, and adaptability.

1. INTRODUCTION

UAVs have found extensive utility across various domains, encompassing applications such as agriculture, surveillance, road maintenance, cultural heritage preservation, and documentation (Ouédraogo et al., 2014). UAV photogrammetry has the capability to produce high-resolution Digital Elevation Models (DEMs), which are recognized as crucial spatial information resources for the exploration of geomorphology and hydrology (Amr et al., 2020). Even though using data from satellites might save money, the pictures they provide are usually not detailed enough to make digital elevation models as precise as those made from methods done on the ground. These methods are also not good for making very detailed maps (Westoby et al., 2012). Additionally, satellite method has limitations in terms of the temporal frequency of its data for economic and social services (Junqing et al., 2012).

Detecting specific structures such as gas stations, airports, and schools is really important for planning smart cities, managing things better, and even for military use (Dell'Acqua, Gamba., 2012). But the old ways of using maps and surveys to find these buildings take a lot of time and work. These methods aren't quick enough to keep up with how fast cities change. Now, because technology has been improved a lot, the pictures we take from above have way more details. This means we can spot different types of buildings using these pictures (Chen, 2007).

In the past, finding specific buildings in pictures involved looking at things like corners, edges, and textures (Cooner et al., 2016). For example, teamwork and others used these features to find Azhar buildings in pictures from above. Similarly, teamwork and others used things like plant data and colour details to find illegal constructions in pictures taken by drones. Even though these methods make sense, they aren't very accurate because they rely on limited information and rules set by people. Also, they don't work well for different types of buildings (Uprety, et al., 2009).

In recent years, object detection methods based on deep learning have made significant breakthroughs for natural images, which can be divided into region- and regression-based methods. Since the breakthrough of the region-based convolutional neural network (R-CNN) (Girshick et al., 2014) for natural images, the combination of a region-based extractor and detection network has become a classic paradigm. In region-based object detection methods, the proposed object box can be generated and then transmitted to the deep convolutional neural network (CNN) for classification and location regression in the second stage. Although the accuracy of methods, such as Faster R-CNN (Ren et al., 2015) and Mask R-CNN (He et al., 2017) are relatively high, they are unable to conform to the requirements of real-time applications. Apart from region-based object detection methods, we have regression-based methods, including You Only Look Once (YOLO) (Redmon et al., 2016), Single Shot Multi-Box Detector (SSD) (Liu et al., 2016), YOLOv2(Redmon, and Farhadi., 2017), and YOLOv3(Redmon, Farhadi., 2018), YOLOv5 (Redmon, Farhadi., 2018) and YOLOv8 (Han et al., 2017) that's we are using.

The rest of this paper is organized as follows. Section 2 describes the study area and details of the model improvements. Section 3 describes the results evaluation indicators, experimental settings and presents the analysis of the experimental results. Section 4 discusses the improved model. Finally, Section 5 concludes this paper.

2. MATERIAL AND METHODS

2.1 Study area and data sources

The study areas chosen were in the campus of Al-Azhar University in Cairo, situated within the Arab Republic of Egypt and in Moscow, Russia. UAV flights were carried out at different altitudes to capture images spanning diverse spatial resolutions and areas of coverage. These flights were executed using a DJI Air 2s drone. The SR were 1-2 centimeter.
UAV flights were conducted at various altitudes to capture images with different spatial resolutions and coverage areas using a DJI Air 2s drone in Cairo, Egypt. Aerial imagery was carried out using DJI Phantom 4 Pro drone in Moscow Russia.
The study incorporated more than 30 GCPs evenly distributed across the surveyed area in every site. RTK-GNSS, utilizing dual-frequency GNSS receivers Trimble R4, were employed to determine the coordinates of these GCPs. Photogrammetric processing of the data was accomplished using Agisoft Metashape Professional and DroneDeploy software. Accuracy assessment of georeferencing was carried out using pre-marked Check Points (CPs) reliably identifiable in the images.

2.2 Methodology

Currently, aerial photography materials obtained from UAVs are being employed to address numerous civil tasks, such as cartography and disaster monitoring. One of the advantages of UAV-based imagery lies in its ability to provide high spatial resolution images, crucial for detailed cartography. A significant application has emerged in mapping using UAV-captured materials. Most UAV platforms are equipped with GNSS receivers based on Micro Electro Mechanical Systems (MEMS) technology, capable of determining spatial coordinates of imaging centers with an accuracy of up to 10 meters during flight.

In this study, a comparison was conducted regarding the geometric accuracy of three different methods of georeferencing, specifically the determination of linear elements in exterior orientation (EO) of images. The first method, known as Direct Georeferencing (DG), primarily relies on the onboard GPS equipment without the use of Ground Control Points (GCPs). The second method, Indirect Georeferencing (IG), relies mainly on GCPs utilized in aerial triangulation (AT). The third method, a modified approach, builds upon the extrapolation of systematic errors in linear EO determination from local route segments to the entire route. It employs Modified Georeferencing (MG) technique and the same three GCPs as the second method.

The study areas chosen were in the campus of Al-Azhar University in Cairo (Figure 1), situated within the Arab Republic of Egypt and in Moscow, Russia. UAV flights were carried out at different altitudes to capture images spanning diverse spatial resolutions and areas of coverage. These flights were executed using a DJI Air 2s drone. The spatial resolutions (SR) were 1-2 centimeter. The study incorporated more than 30 evenly distributed GCPs across the surveyed area. RTK-GNSS, utilizing dual-frequency GNSS receivers Trimble R4, were employed to determine the coordinates of these GCPs. Photogrammetric processing of the data was accomplished using Agisoft Metashape Professional and DroneDeploy software. Accuracy assessment was carried out using pre-marked Check Points (CPs) reliably identifiable in the images.

Figure 1. Study area, Al-Azhar University, Cairo, Egypt.

The three methods compared are direct georeferencing, relying mainly on the onboard GNSS equipment without the use of GCPs; indirect georeferencing, primarily reliant on three GCPs used in aerial triangulation; and the modified technique of georeferencing, which extrapolates systematic errors in linear EO determination from local route segments to the entire route, employing a Linear Relationship (LR) model and the same GCPs as the second method (Elsheshtawy et al., 2202).

The proposed MG technique for generating EO (XNi, YNi, ZNi) obtained by GNSS on the drone platform and use it in georeferencing, consists of the following steps:
• Obtaining EO (XGi, YGi, ZGi) for all images (n) through on-board drone GNSS (Direct Georeferencing).
• Computing EO (XRi, YRi, ZRi) for that limited images (k) which have in it the three GCPs locally located in a circle about two meters diameter (Indirect Georeferencing with limited images. k << n.
• The computation of discrepancies (DXi, DYi, DZi) in the Exterior Orientation (EO) derived from the initial and subsequent steps, limited to a certain number of images (k), follows formulas similar to equation (1):

$$DXi = XGi - XRi \quad (1)$$

Where:
DXi represents the disparity between XGi and XRi (DXi = XGi - XRi).
XGi denotes the camera station position X of image i attained through drone GPS (X GPS).
XRi signifies the camera station position X of image i as determined by the initial georeferencing (X reference).
n corresponds to the total number of images considered.
• Computing the average value of differences DXi (mean)obtained from formulas (1) for k images using formulas like (2):

$$mean = (\sum 1\_(i=1) \text{ ^k } DXi) / k \quad (2)$$

Taking the value of mean as the systematic error XGi.
• Calculating generated EO (XNi, YNi, ZNi) for all remaining images using formulas like (3):

$$XNi = XGi - mean \quad (3)$$

Where:
XNi represents the newly generated sensor location X for image i in the remaining set of images (X new).

XGi signifies the camera station position X of image i acquired through drone GNSS (X GNSS).• Similar equations are used for the Y and Z coordinates of imaging centers.
• The generated EO (XNi, YNi, ZNi) is used in final georeferencing.

Another aspect of this research is dedicated to addressing Lane detection techniques, which fall into two primary categories: traditional machine vision methods and modern deep learning approaches. Traditional methods rely on statistical analysis of image features, such as colour, gray levels, and edges, while deep learning techniques employ convolutional neural networks for robust feature extraction. While traditional methods offer acceptable accuracy levels, they often entail intricate processes with substantial human involvement, limiting their practical applicability. On the other hand, deep learning solutions, exemplified by YOLO, strike a balance between accuracy and detection speed.

Against this backdrop, this paper presents a novel lane detection system based on an enhanced YOLOv3 architecture. The proposed model leverages YOLOv3's rapid detection capabilities while addressing inherent limitations. The contributions of this study include a grid-based image division to accommodate varying lane line densities, optimization of detection scales for improved small target detection, adoption of a Darknet-49 backbone architecture, and parameter refinement for focused lane detection. This research aims to amplify lane detection technology by synergizing the strengths of YOLOv3 with tailored enhancements, promising heightened robustness and operational efficiency.

The process of lane detection and recognition comprises the following components: image acquisition, image preprocessing, image segmentation, and edge detection, feature point identification, and lane line recognition, as shown in Figure 2.

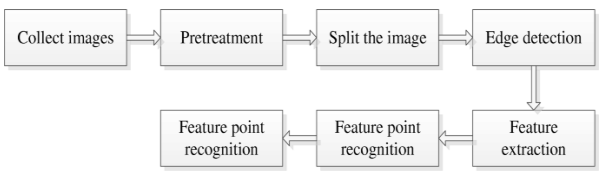

Figure. 2 Method to detect lane lines.

2.2.1 Road image acquisition and pretreatment:

This discussion centers on the significance of image pre-processing in the context of lane line detection UAVs.

The study utilizes a Canon EOS 100D DSLR camera for imagery acquisition within a smart car identification and navigation system. Real-world road conditions introduce various external factors, resulting in artifacts like spots and pits in UAV-captured images. Such interference, coupled with potential image quality degradation, directly impacts the precise detection and recognition of lane line information. To address this, pre-processing of collected images is crucial, involving interference reduction, target information enhancement, streamlined image processing, and heightened detection accuracy.

Image capture occurs amidst dynamic road scenes, often at high speeds. Stringent safety considerations dictate specific capture conditions, such as during red light intervals or at pedestrian crossings. Image focus is optimized for UAV operations, ensuring swift and accurate capture to maintain image quality.
Images from UAV cameras, influenced by real-world conditions, may contain interference factors like spots and depressions. This degradation impairs accurate lane line detection. Image pre-processing aims to remove noise, enhance target details, and streamline algorithms, encompassing greying, noise reduction, and image enhancement techniques.

Greying reduces colour complexity, aiding algorithmic efficiency. Leveraging wavelet decomposition improves frequency resolution, while image enhancement emphasizes region-of-interest characteristics, reinforcing grayscale attributes and lane line information.
UAV-captured images undergo processing to expedite detection algorithms. Grayscale conversion simplifies processing, while addressing salt and pepper noise, the most common noise type. Median filtering emerges as effective in noise reduction, preserving image quality and contour information. Its real-time performance makes it an optimal choice for noise mitigation

By leveraging a streamlined dataset and custom pre-processing, we prepared the model for training on CUDA, ensuring rapid processing and sufficient memory allocation. The customized YOLOv8 nano model demonstrated impressive accuracy in identifying buildings from satellite images, thereby validating its effectiveness in addressing this specialized task.
Our approach to achieving accurate building detection through the YOLOv8 pre-trained model is marked by a comprehensive and systematic methodology:
Data Gathering: We collected a concise building detection dataset from Kaggle, comprising approximately 160 images categorized into training (~140 images), validation (~10 images), and testing (~10 images) sets, accompanied by their corresponding labels.

The Massachusetts Buildings Dataset consists of 151 aerial images of the Boston area, with each of the images being 1500 × 1500 pixels for an area of 2.25 square kilometers. Therefore, the complete dataset encompasses approximately 340 square kilometers, as depicted in Figure 3. This dataset is partitioned into a training set comprising 137 images, a test set containing 10 images, and a validation set encompassing 4 images.
he desired maps were derived by converting building footprints, acquired from the OpenStreetMap project, into raster format. The dataset was limited to areas with an average omission noise level of approximately 5% or lower. The substantial volume of building footprint data of excellent quality was feasible to gather due to the contribution of building footprints for the entire city by the City of Boston to the OpenStreetMap project.
The dataset predominantly encompasses urban and suburban regions, encompassing buildings of varying sizes, including individual houses and garages, within the labelling. The datasets utilize imagery made available by the state of Massachusetts, all of which has been rescaled to a resolution of 1 pixel per square meter. The target maps within the dataset were generated through the utilization of data sourced from the OpenStreetMap project.
Target maps allocated for the test and validation segments of the dataset underwent manual correction to enhance the precision of evaluations. Figure 3. Image from training of dataset. 2.2 Dataset Preparation Data Pre-processing: The labels were initially available as Black-White images, but YOLO necessitates annotation in the form of boundary boxes provided in text files. Thus, we developed Python scripts that operate by taking these "BW images" labels as shown in Figure 4. along with the original images.

Figure 3. Image from training of dataset.

2.2.1 Dataset Preparation

Data Pre-processing: The labels were initially available as Black-White images, but YOLO necessitates annotation in the form of boundary boxes provided in text files. Thus, we developed Python scripts that operate by taking these "BW images" labels as shown in Figure 4. along with the original images. The scripts then proceed to create bounding boxes around each building in the corresponding original image. These images with superimposed boundary boxes are saved as new images as shown in Figure 5. and the associated data for each bounding box, in the format "class_id center_x center_y width height," is extracted and stored in annotation files
The models were executed using CUDA to enhance processing speed and enable the allocation of greater memory capacity. A YOLOv8 model of nano size was trained on these customized datasets, specifically tailored to facilitate the detection of buildings in satellite imagery. The results of this training exhibited notably high accuracy across both the training and validation phases.

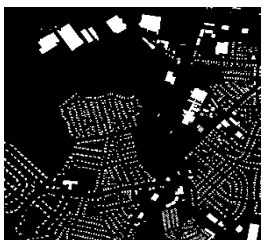

Figure 4. Image from training of dataset.

Figure 5. Detecting buildings from satellite image.

Model Training: To optimize performance, we utilized CUDA to enhance the model's processing speed and memory allocation. We selected a YOLOv8 nano model tailored to our custom dataset's specifications. This model was specifically designed for detecting buildings in satellite imagery as shown in Figure 6

Figure 6. Detect building from dataset.

Accuracy and Performance: Throughout both the training and validation stages, the trained YOLOv8 nano model displayed remarkable accuracy as shown in Figure 7. Its specialized architecture, focusing on identifying buildings in satellite images, delivered promising outcomes.

Figure 7. YOLOv8 Ultralytics: State-of-the-Art YOLO Models.

2.3 Experimental Settings

2.3.1 Evaluation Indicators: To provide a quantitative assessment of the chosen models' performance, we employed the average precision (AP) and the Precision Recall Curve (PRC). Furthermore, the F1 score (Tian et al.,2019) and Frames Per Second (FPS) were employed to gauge the model's effectiveness and detection speed.

2.3.1.1 Precision Recall Curve: The Precision Recall Curve (PRC) is depicted with precision on the Y-axis and recall on the X-axis. Prior to constructing the PRC, the calculation of precision and recall (Benjdira et al., 2019) is imperative. The equations for the precision, P, and recall rate, R **(Figure 8),** are as follows:

$$P = \frac{TP}{FP + TP'} \quad (4)$$

$$R = \frac{TP}{FN + TP'} \quad (5)$$

Where:

TP is the number of correctly detected positive samples.

FP is the number of negative samples detected by error as positive samples.

FN is the number of positive samples not detected.

If the area overlap ratio between the predicted bounding box and ground-truth bounding box is larger than 0.5, we set the

Alternatively, if no overlap exists, it is designated as a false positive (FP). Furthermore, in cases where multiple predicted bounding boxes intersect with the same ground-truth bounding box, only the box with the highest overlap is deemed a true positive (TP). It's important to note that precision and recall rate possess an inverse correlation. Refer to Figure 8 for the confusion matrix representing predicted outcomes and ground truth.

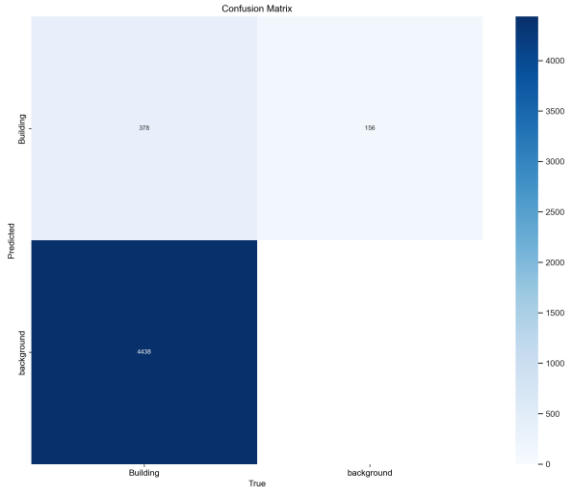

Figure 8. Confusion matrix for predicted results and ground truth.

3. RESULTS

3.1 Testing the Modified Technique of Georeferencing

The test demonstrating a Root Mean Square Error (RMSE) of 2-3 cm was selected as the benchmark for subsequent tests. The calculated sensor location obtained through this reference test was considered reliable data. To evaluate the accuracy of the sensor location generated by the Modified technique, the disparities in the X, Y, and Z coordinates were computed between the generated data and the reference data of the sensor location for each image figure 9. The same approach was taken for calculating the differences between the drone's GNSS data and the reference sensor location data figure 10.

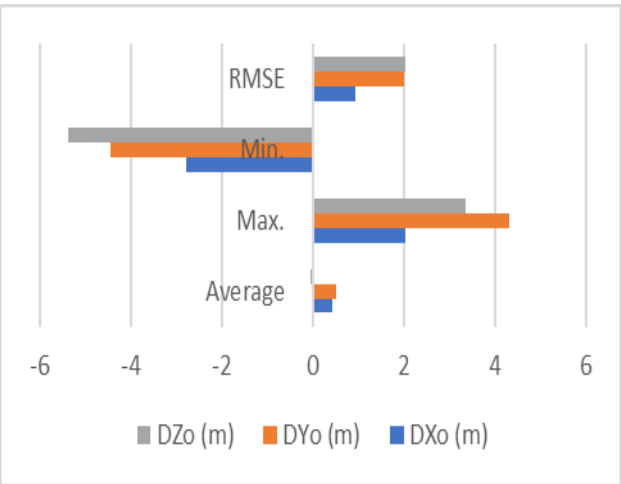

Figure 9. Descriptive statistics of the differences the Modified technique and considered reliable data.

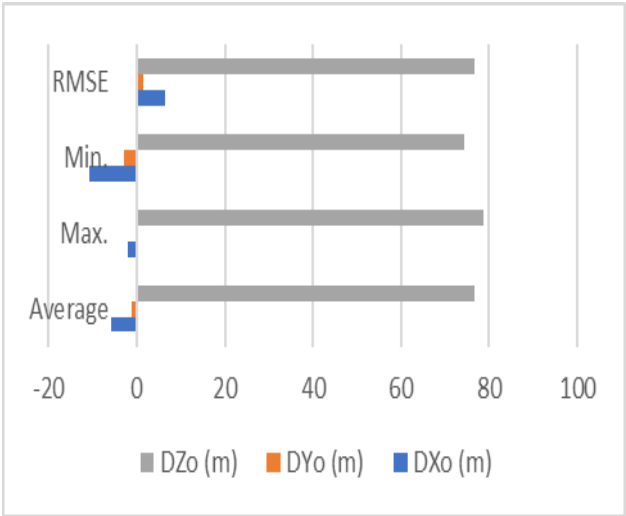

Figure 10. Descriptive statistics of the differences drone GNSS and considered reliable data.

Based on the preceding outcomes, it becomes evident that the sensor locations generated through the modified georeferencing technique exhibit greater accuracy compared to the utilization of drone GNSS data in direct georeferencing.

3.2 The PRC Evaluation

In the context of the object detection method, the Precision-Recall Curve serves as a fundamental metric for gauging both the robustness and efficacy of the approach. Upon examining the curve, it becomes evident that as the recall rate increases, there is a gradual decline in precision (Figure 11).

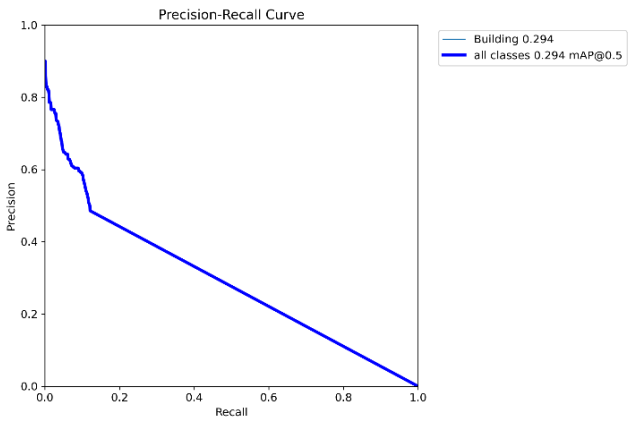

Figure 11. PR curve.

The endeavour yielded intriguing results. The model exhibited remarkable accuracy in detecting buildings from satellite images, affirming its initial proficiency. However, the true test lay in its performance with drone-captured images. While improvements were achieved, the model's struggles persisted, particularly with close-up shots. The misidentification of pillars and air conditioners as buildings remained a challenge. The augmented dataset and fine-tuning will lead to a progress as we'll train our model on drone images also, but further refinement is essential to bridge the disparity between the two types of images.

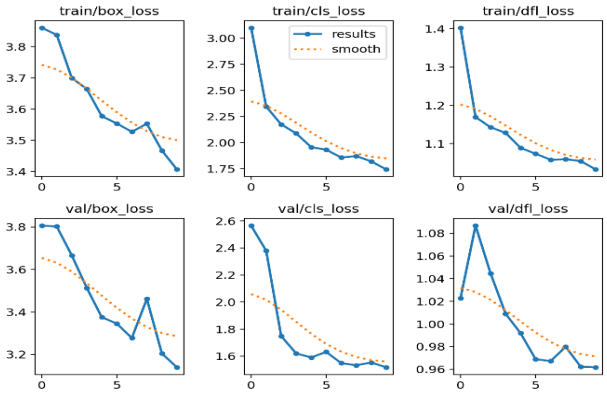

Figure 12. Train and validation loss.

4. DISCUSSION

Numerous studies have been conducted to utilize remote sensing images for extracting information about collapsed or damaged buildings following the 2008 Wenchuan earthquake and the 2010 Yushu earthquake. An object-oriented change detection method was applied employing multiple classifiers to identify building damage after the Yushu earthquake. This involved integrating multiple feature extraction, selection, and random subspace recognition techniques to enhance the multi-classifier system, leading to an overall accuracy of 88.45% (Zhao et al., 2018). The researchers employed LiDAR data and high-resolution Quickbird remote sensing data in the Yushu disaster zone. They combined object-oriented classification with SVM technology to extract information about collapsed buildings, achieving a total extraction accuracy of 82.21% (Wen. Ji et al., 2015). adopted a method that involved using building vector data to extract individual building objects from remote sensing images. They then employed CNN for classifying fully collapsed buildings and those that were intact or minimally affected, attaining an average accuracy of 78.6%. While most of these studies achieved high accuracy, they often relied on a variety of data sources such as pre-earthquake remote sensing images, LiDAR data, and building vector data. Obtaining such data immediately post-earthquake can be challenging. In contrast, our study employed the CNN-based object detection method YOLOv8 to detect collapsed buildings, yielding a notable accuracy of 90.89%. Our approach proves more practical as YOLOv8 exclusively requires post-earthquake remote sensing images, eliminating the need for pre-earthquake data. Furthermore, YOLOv8 can detect collapsed buildings without the aid of building vector data, addressing practical limitations posed by the latter.

To mitigate overfitting during CNN training, we undertook specific measures. The limited seismic data collection, coupled with the necessity to identify individual collapsed buildings using high-resolution data, led to a small dataset. This could result in overfitting when training a large convolutional neural network on the training set. To counter this, we expanded and enriched the dataset to enhance sample diversity. Additionally, in the YOLOv8 network structure, replacing Darknet53 with the lightweight CNN ShuffleNet v2 reduced the parameters and effectively alleviated overfitting. In instances where the loss function value stagnated after 50 epochs during training, we terminated training prematurely to prevent excessive learning.

In the context of enhancing detection precision for damaged buildings, modifications were introduced to the YOLOv8 loss function. The original cross-entropy loss function employed in YOLOv8 for predicting center point coordinates and box dimensions was substituted with the GIoU loss. The GIoU loss was adopted as it accurately describes the relationship between prediction and true boxes, allowing for accurate evaluation of loss during the training process.

5. CONCLUSION

The experimental work in this study demonstrates that the proposed modified georeferencing technique to refine sensor locations, is more accurate than the drone GNSS data, all without incurring additional costs. This is particularly significant in projects where GCPs cannot be distributed extensively or where is bad GNSS signal places. Furthermore, this study has contributed to the realm of road monitoring and navigation through innovative lane detection techniques. The integration of image pre-processing, YOLO-based edge detection, and lane line feature point recognition has yielded a robust framework. Image pre-processing effectively elevates lane detection accuracy by minimizing noise and refining target information. YOLO's real-time edge detection capabilities excel in identifying and localizing lane boundaries, even in complex scenarios.

Amidst the ever-evolving landscape of AI and UAV technology, this paper delves into both the potential and intricacies of implementing advanced AI models, such as YOLOv8, for building detection using UAV and satellite images. While the model displayed its capability in analyzing satellite images, the complexities of drone imagery have posed challenges that necessitate ongoing exploration. The path ahead involves a steadfast commitment to innovation, method refinement, and recognition of the nuances involved in close-range image analysis. As researchers continue to unveil the synergy between AI and UAV imagery, we edge closer to a future where structures are not just understood from a distance but also up close, heralding a new era of precise, adaptable, and comprehensive building detection.

Furthermore, this study has contributed to the realm of road monitoring and navigation through innovative lane detection techniques. The integration of image pre-processing, YOLO-based edge detection, and lane line feature point recognition has yielded a robust framework. Image pre-processing effectively elevates lane detection accuracy by minimizing noise and refining target information. YOLO's real-time edge detection capabilities excel in identifying and localizing lane boundaries, even in complex scenarios.

REFERENCES

Amr E. M., Larisa G. A., Mohamed E. A.: Low-Cost Technique of Enhancement Georeferencing for UAV Linear Projects. *In Proceedings of the 2020 3rd International Conference on Geoinformatics and Data Analysis* 2020 Apr 15 (pp. 76-80).

Benjdira, B.; Khursheed, T.; Koubaa,: A. Car detection using unmanned aerial vehicles: Comparison between Faster R-CNN and YOLOv3. *In Proceedings of the International Conference on Unmanned Vehicle Systems-Oman (UVS), Sultan Qaboos Univ, Muscat, Oman*, 5–7 February 2019; pp. 1–6.

Chen, W.: Research of Remote Sensing Application Technology Based on Earthquake Disaster Assessment; China Earthquake Administration Lanzhou Institute of Seismology: Lanzhou, China, 2007.

Cooner, A.; Shao, Y.; Campbell, J.: Detection of urban damage using remote sensing and machine learning algorithms:

Revisiting the 2010 Haiti earthquake. Remote Sens. 2016, 8, 868.

Elsheshtawy AM, Gavrilova LA.: Improving Linear Projects Georeferencing to Create Digital Models Using UAV Imagery. *InE3S Web of Conferences* 2021 (Vol. 310, p. 04001). EDP Sciences.

Dell'Acqua F, Gamba P.: Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE*. 2012 Jul 10;100(10):2876-90.

Elsheshtawy AM, Gavrilova LA.: Improving Linear Projects Georeferencing to Create Digital Models Using UAV Imagery. *InE3S Web of Conferences* 2021 (Vol. 310, p. 04001). EDP Sciences.

Girshick, R.; Donahue, J.; Darrell, T.: Rich feature hierarchies for accurate object detection and semantic segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, *OH, USA*, 23–28 June 2014; pp. 580–587.

Han, X.; Zhong, Y.; Zhang, L.: An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery. *Remote Sens*. 2017, 9, 666.

He, K.; Gkioxari, G.; Dollár, P. Mask R-CNN.: *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA*, 21–26 July 2017; pp. 2961–2969.

Junqing C, Zongjian L, Xiaojing W, Yongrong L.: Application of UAV system for low altitude photogrammetry in Shanxi. The *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2012 Jul 24;39:351-4.

Liu, W.; Anguelov, D.; Erhan, D. SSD: Single shot multibox detector. *In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands*, 8–16 October 2016; pp. 21–37.

Ouédraogo MM, Degré A, Debouche C, Lisein J.: The evaluation of unmanned aerial system-based photogrammetry and terrestrial laser scanning to generate DEMs of agricultural watersheds. *Geomorphology*. 2014 Jun 1;214:339-55.

Redmon J, Farhadi A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018 Apr 8.

Redmon, J.; Divvala, S.; Girshick, R.: You only look once: Unified, real-time object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA*, 27–30 June 2016; pp. 779–788.

Redmon, J.; Farhadi, A.: YOLO9000: Better, faster, stronger. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

Ren, S.; He, K.; Girshick, R. Faster R-CNN: Towards real-time object detection with region proposal networks. *In Proceedings of the Annual Conference on Neural Information Processing Systems, Montreal, QC, Canada*, 7–12 December 2015; pp. 91–99.

Tian Y, Yang G, Wang Z, Wang H, Li E, Liang Z.: Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and electronics in agriculture*. 2019 Feb 1;157:417-26.

Uprety P, Yamazaki F, Dell'Acqua F.: Damage detection using high-resolution SAR imagery in the 2009 L'Aquila, *Italy, earthquake. Earthquake Spectra*. 2013 Nov 1;29(4):1521-35.

Wen, X.; Bi, X.; Xiang, W.: Object-oriented collapsed building extraction from multi-source remote sensing imagery based on SVM. North China Earthq. Sci. 2015, 33, 13–19.

Westoby MJ, Brasington J, Glasser NF, Hambrey MJ, Reynolds JM.: 'Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology. 2012 Dec 15;179:300-14.

Zhao, Y.; Ren, H.; Cao, D.: The research of building earthquake damage object-oriented change detection based on ensemble classifier with remote sensing image. *In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Valencia, Spain*, 22–27 July 2018; pp. 4950–4953.