2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

YOLO: A Competitive Analysis of Modern Object Detection Algorithms for Road Defects Detection Using Drone Images

***Amit Hasan Sadhin**[1], **Siti Zaiton Mohd Hashim**[1], **Hussein Samma**[2], **Nurulaqilla Khamis**[3]

[1]Faculty of Computing, University of Technology Malaysia, Johor, Malaysia.
[2]SDAIA-KFUPM Joint Research Center for Artificial Intelligence, KFUPM, Saudi Arabia.
[3]Faculty of Electrical Engineering, University of Technology Malaysia, Johor, Malaysia.
*Corresponding Author.

Received 04/05/2023, Revised 10/10/2023, Accepted 10/10/2023, Published Online First 20/11/2023, Published 1/6/2024

Abstract

Efficient identification of road defects is a critical concern for road safety and infrastructure upkeep. This research employs drone-captured imagery and advanced object detection algorithms to expedite defect recognition, with a specific focus on determining the optimal algorithm for prompt and precise detection. The importance of timely road defect detection, crucial for mitigating potential hazards, remains central. A comprehensive comparative analysis of contemporary object detection algorithms, encompassing YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and YOLOv7. The results of this study highlight YOLOv7 as the most efficient, with a notable mAP of 68.3%, closely followed by YOLOv5l (66.8%), YOLOv5m (66.3%), YOLOv5x (66%), and YOLOv5s (63%). The integration of drone-derived imagery, capturing distinct gradients, significantly enhances defect detection accuracy. Beyond road safety, this study offers valuable insights to computer vision and machine learning practitioners. By bridging technological innovation with practical implementation, it holds potential to advance road safety and transportation infrastructure quality and the use of revolutionary drone technology.

Keywords: Convolution Neural Network, YOLO, drone images, CSPDarknet, Input Size.

Introduction

The march of human civilization has ushered in profound advancements in science and technology. However, concomitant with this progress are vulnerabilities that have become more pronounced, particularly concerning safety and human lives. Among these concerns, road transportation safety stands as a critical focal point, given the significant repercussions of deteriorating roads on individual lives and the global economy. Roads serve as conduits to diverse regions, granting access to employment, social amenities, healthcare, and education, thereby fostering economic development and poverty alleviation. To address the escalating demands posed by burgeoning vehicular traffic, countries are constructing diverse road types to establish efficient and secure connectivity.

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Notwithstanding these efforts, road degradation is an ongoing challenge attributed to factors like traffic congestion, suboptimal design and materials, and inadequate maintenance. The consequence of such degradation is the emergence of prevalent road cracks and potholes, culminating in numerous accidents annually and impeding societal and economic progress. Consequently, road maintenance assumes paramount importance, with road inspection serving as a pivotal precursor to effective upkeep. Traditional human-based inspections have been the norm, yet recent strides in drone technology offer a compelling alternative for inspecting critical safety systems. Drones expedite data collection, enhancing inspection efficiency, accuracy, and cost-effectiveness. Furthermore, drone deployment mitigates personnel safety concerns, particularly in high-risk environments such as highways. Notably, studies have underscored the efficacy of drones in inspecting infrastructure like bridges [1].

Prior research has explored Deep Learning Algorithms for road crack identification. The YOLO (You Only Look Once) family of networks has emerged as a noteworthy contender, demonstrating remarkable accuracy in object detection under critical contexts. Variants like YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and YOLOv7 exhibit potential for detecting objects against complex backgrounds, including roads. However, a pivotal question persists: which YOLO algorithm version is optimal for detecting road cracks and potholes? This study embarks on comparative experiments, aiming to ascertain the most reliable YOLO version for road defect detection via drone imagery.

This paper focuses on the gaps and opportunities in the current situation. It discusses the challenges encounter by the researchers, explains how our proposed algorithms can overcome these challenges, highlights what this research contributes, and gives an overview of how the paper is organized.

Literature Review:

A nation's progress and development hinge critically on its road networks, serving as indispensable conduits that not only facilitate physical infrastructure but also bestow vital social advantages. These interconnected pathways are pivotal for a country's advancement and play an instrumental role in addressing poverty by enabling access to essential daily necessities and fostering enhanced communication channels. For instance, the European road network, encompassing a vast expanse of 5.5 million kilometers, stands as a testament to this significance, commanding a staggering valuation exceeding €8,000 billion. Cognizant of these imperatives, a substantial allocation of over $400 billion is annually directed toward road maintenance. However, the protracted processes entailed in pinpointing road-related issues and subsequent maintenance operations impose escalating costs each year. The persistently mounting expenses are further exacerbated by road afflictions such as cracks and potholes, perpetuating a cycle of hundreds of thousands of global accidents annually. This grim toll not only exacts a devastating loss of lives but also casts a pall over the broader global economic dividends. In light of these challenges, there arises an imperative need to ameliorate the efficiency of road problem identification and maintenance processes, thereby curtailing both human and economic tolls.

Road Cracks and Potholes:

Road cracks and potholes are frequent occurrences in pavement, highways, and roads, and they can seriously affect both road safety and maintenance expenses. The most typical types of road cracks are alligator crack, longitudinal, and transverse as shown in Fig. 1. Transverse fractures are typically induced by fatigue or heat forces, whereas longitudinal cracks are typically caused by shrinkage. Combining the two, alligator cracking is typically brought on by high traffic volumes or continuous loading. On the other hand, freeze-thaw cycles and water infiltration are typically to blame for potholes. They can range in size and depth and pose a severe risk to both motorists and pedestrians.

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

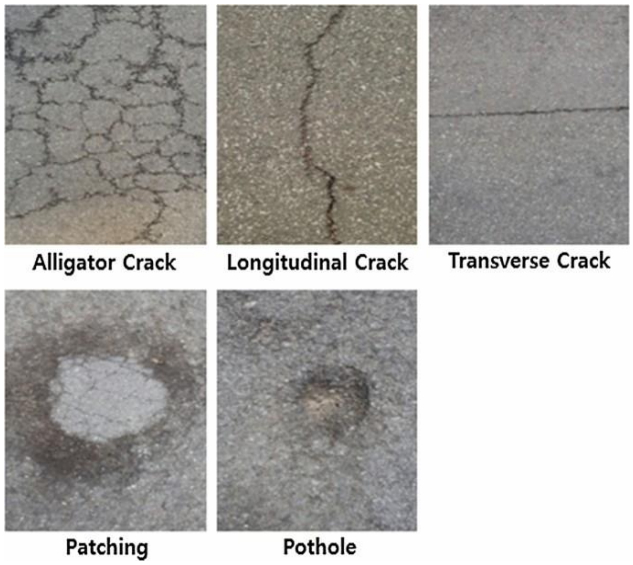

Figure 1. Different types of road cracks [2]

Moreover, road cracks and potholes are major factors in accidents and fatalities and can seriously affect traffic safety. Around one-third of all traffic fatalities in the US are caused by road conditions like potholes, cracks, and ruts, according to Federal Highway Administration (FHWA) research from 2019. Road cracks and potholes increase the risk of accidents and can also harm vehicles, lengthen travel times, and cost local governments more money to maintain. Poor road conditions can also hurt the economy. Businesses may be less likely to invest in places with inadequate infrastructure [3].

Drone Inspection:

Drones have revolutionized the field of inspection by providing efficient and cost-effective solutions. The ability to inspect hard-to-reach or hazardous areas without risking human lives makes drones an attractive option for inspecting bridges, buildings, and roads as highlighted in Fig. 2. Drones provided high-resolution images and video footage of the bridge, allowing for more detailed inspection compared to traditional methods. Additionally, drones were able to inspect hard-to-reach areas of bridges, such as the undersides of the deck, which are difficult to access with conventional inspection methods. Likewise, drones are being used for powerline inspection, where drones are able to detect defects in power lines with a top-notch accuracy and efficiency, reducing the time and cost required for inspections. Drones were able to inspect power lines in hazardous or difficult-to-reach locations, improving safety for inspectors. Moreover, for leaf disease and more, drones are being used in agriculture to identify nitrogen-deficient crops [4].

Figure 2. Pavement inspection using drone [5]

YOLO Algorithms in Road Inspections:

Different methods have been used for detecting objects, such as Fourier and Wavelet transformations, to reduce computational complexity and enhance moving object detection [6]. YOLO versions are state-of-the-art algorithm in object detection. It has gained popularity due to its speed and accuracy. In recent times, researchers focused on the application of YOLO algorithms for road cracks and potholes detection.

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

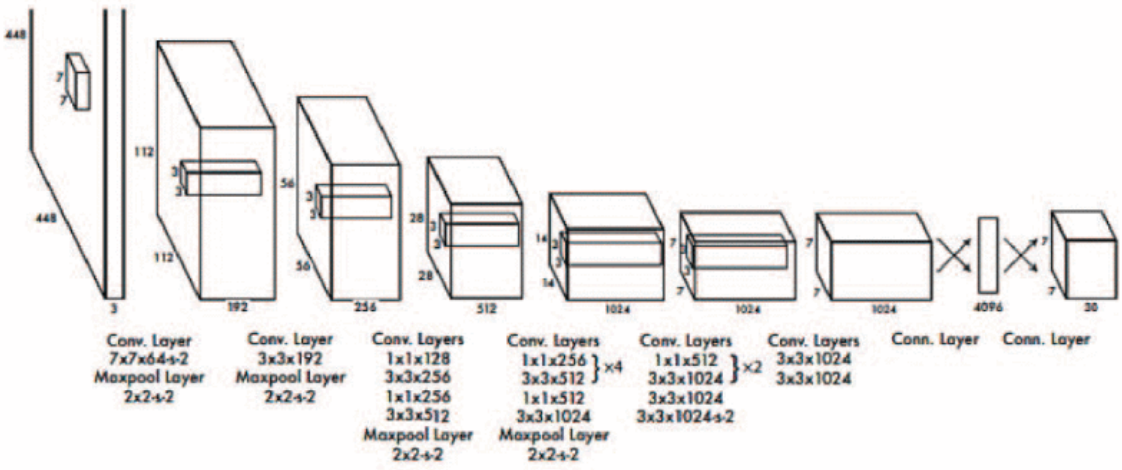

Figure 3. Yolo architecture [7]

All the YOLO version follows the same pattern as in Fig. 3, YOLOv2 outperforms the other state-of-the-art algorithms by detecting road cracks and potholes with more speed and accuracy. Similarly, YOLOv3 produced high accuracy and high speed in small datasets. For example, YOLOv5 demonstrated a remarkable balance between accuracy and computational efficiency, achieving an impressive mAP value of 63.54% while maintaining a swift inference time of approximately 0.61 seconds per image[8]**.** The use of anchor boxes present in the YOLO network helps predict the object's shape and size within the bounding boxes. Different YOLO versions show different numbers of layers, such as the YOLOv3 architecture consisting of 53 convolutional layers [9]. The basic YOLO architecture is made up of a single CNN that takes the full image and returns a fixed number of bounding boxes and class probabilities. The input image is divided into cells, and each cell predicts a set number of bounding boxes, as well as the confidence value and class probabilities for each box [10]. Several convolutional layers are followed by max-pooling layers in the architecture, which reduces the spatial resolution of the feature maps while increasing the number of feature maps. Each bounding box's confidence score is the product of the chance that the box includes an object and the probability that the box is precisely localized. After that, non-maximum suppression is employed to eliminate overlapping boxes with low confidence scores. The YOLO architecture has undergone several versions, with each new version introducing new features and achieving higher accuracy on object detection tasks.

Proposed Work:

Dataset:

The dataset does contain a total of 1000 images of road cracks and potholes, which are collected using a DJI spark drone (shown in Fig. 4). The images were captured from the different roads of the University of Technology Malaysia. The images were taken at the height of 25 to 35ft. Every image in the dataset is unique and has never been used in any research before. There is no duplicate image in the dataset, except the same cracks and potholes images were taken at different angles and different heights. There are some images where multiple cracks and potholes can be seen, as well as single cracks and potholes. The images were in different real-time scenarios and under various circumstances.

Figure 4. DJI Spark drone

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

The collected drone images are used for training the YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, YOLOv6 and YOLOv7 models. The images were annotated using LabelImg, where the images were selected in size of 416x416.

Dataset Description:

The images incorporate a total of 2 classes which are cracks and potholes. Some of the images contain multiple cracks and potholes as shown in Fig. 5. The images were separated into three groups: 80% for training, 10% for testing, and 10% for validation. The size and resolution of the training image datasets have an impact on the performance of DL models[11].

Table 1. Collected datasets and different class elements

Dataset	Training	Testing	Validation	UTM Roads	Total images
Road Cracks	575	71	71	717	1000
Potholes	393	48	51	492	

Figure 5. Images captured using drone

Dataset Augmentation:

Dataset augmentation is a fundamental process to extend the training sets. Therefore, various augmentations are included in data pre-processing, such as flipping the images vertically and horizontally and applying rotation and scaling to improve the images. Via data augmentation, the dataset has been tripled the training sets and made a new version than the actual one to train the model with different inclination images.

Table 2. Dataset after augmentation

Dataset	Training	Testing	Validation	Total images
Cracks & Potholes	2397	101	99	1000

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Method:

In recent times, significant strides have been witnessed in the realm of object detection through deep learning techniques. These approaches can be broadly categorized into two main types: dual-stage detection and single-stage detection. Single-stage object detection has emerged as a preferred choice for scenarios demanding rapid processing and real-time applications like object tracking. Within the ambit of two-stage object detection techniques, exemplified by R-CNN, Faster R-CNN, and Mask R-CNN, a twofold procedure is adopted: object proposal stage and object classification. Initially, a set of object proposals or regions of interest (RoIs) is generated in the first stage using a region proposal network (RPN) or selective search. Subsequently, these proposals are harnessed for object classification and bounding box refinement in the second stage.

Conversely, single-stage object detection methodologies like YOLO, SSD, and RetinaNet operate within a solitary step to detect objects. These methods directly predict class labels and bounding box coordinates for all objects within the image, circumventing the need for a distinct region proposal phase. Typically, single-stage detectors exhibit higher processing speed compared to their dual-stage counterparts. Striving for a balance between accuracy and efficiency, hybrid object detection algorithms such as the EfficientDet series amalgamate both single-stage and two-stage techniques through a compounded scaling approach. This fusion contributes to the attainment of superior detection outcomes while maintaining computational efficiency.

YOLO:

Redmon et al., YOLO (You Only Look Once) algorithm is one of the most advanced single-stage object identification techniques. YOLO is well-known for its simplicity, speed, and precision[13] . As in the Fig. 6, the YOLO algorithm divides the input image into cells and predicts class probabilities and bounding boxes for each one. To extract features, the algorithm first applies a deep convolutional neural network (CNN) to the input image. The resulting feature map is then divided into a grid of cells, with each cell responsible for detecting items that fall within its boundaries.

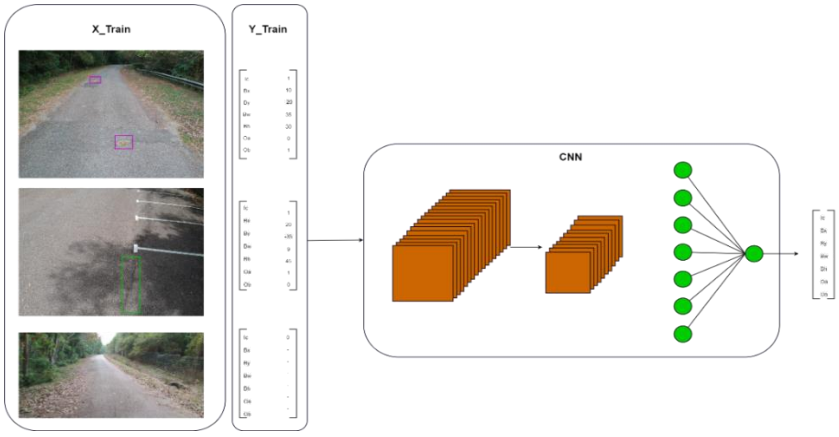

Figure 6. Deep Convolutional Neural Networks approach for YOLO algorithms

For each cell, the procedure involves forecasting class probabilities and delineating bounding boxes for a predefined set of anchor boxes. These anchor boxes constitute pre-established shapes and dimensions, instrumental in projecting object location and dimensions within the cell. These class probabilities and bounding box forecasts are subsequently amalgamated to generate the ultimate detection outcomes. Incorporating the non-maximum suppression (NMS) technique, the YOLO algorithm filters out redundant detections. NMS evaluates the projected bounding boxes, removing those displaying substantial overlap with other predicted boxes. A noteworthy attribute of the YOLO algorithm is its capacity to execute object detection in a single pass. This efficiency empowers real-time processing of images using conventional GPUs. Remarkably, YOLO is adept at detecting diminutive objects, a challenge that often perplexes other prevalent object detection techniques.

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

In this study, YOLOv5 versions and YOLOv7 were taken for the competitive analysis of the capabilities of these two models among each different version and the comparison with the state-of-the-art object detection algorithm. YOLOv5 is chosen due to its speed, accuracy and versatile behaviors, whereas YOLOv7 is a recent stable version among YOLO family networks.

YOLOv5:

The architecture of YOLOv5 consists of a few main components that work together to perform object detection on input images. Fig. 7, highlighted the detail view of the used YOLO architecture.

Backbone: In YOLOv5, the foundational architecture employs a CSPDarknet-53 network as its backbone, depicted in Fig. 7 below. This CSPDarknet-53 network is a derivative of the Darknet neural network structure. Comprising convolutional layers, residual blocks, and down-sampling layers, the CSPDarknet-53 network adeptly extracts salient features from the input image.

Neck: Within the YOLOv5 framework, the neck network takes the form of a feature pyramid network (FPN). This FPN leverages the features obtained from the underlying backbone network and generates a feature pyramid encompassing diverse feature sizes. This strategic approach enhances the algorithm's ability to detect objects of varying dimensions, thereby augmenting the precision of object detection.

Head: In the context of YOLOv5, the head network assumes responsibility for foreseeing both class probabilities and bounding boxes pertaining to every object within the input image. This component comprises convolutional layers along with predefined anchor boxes, which serve as predetermined templates used to anticipate object positions and dimensions.

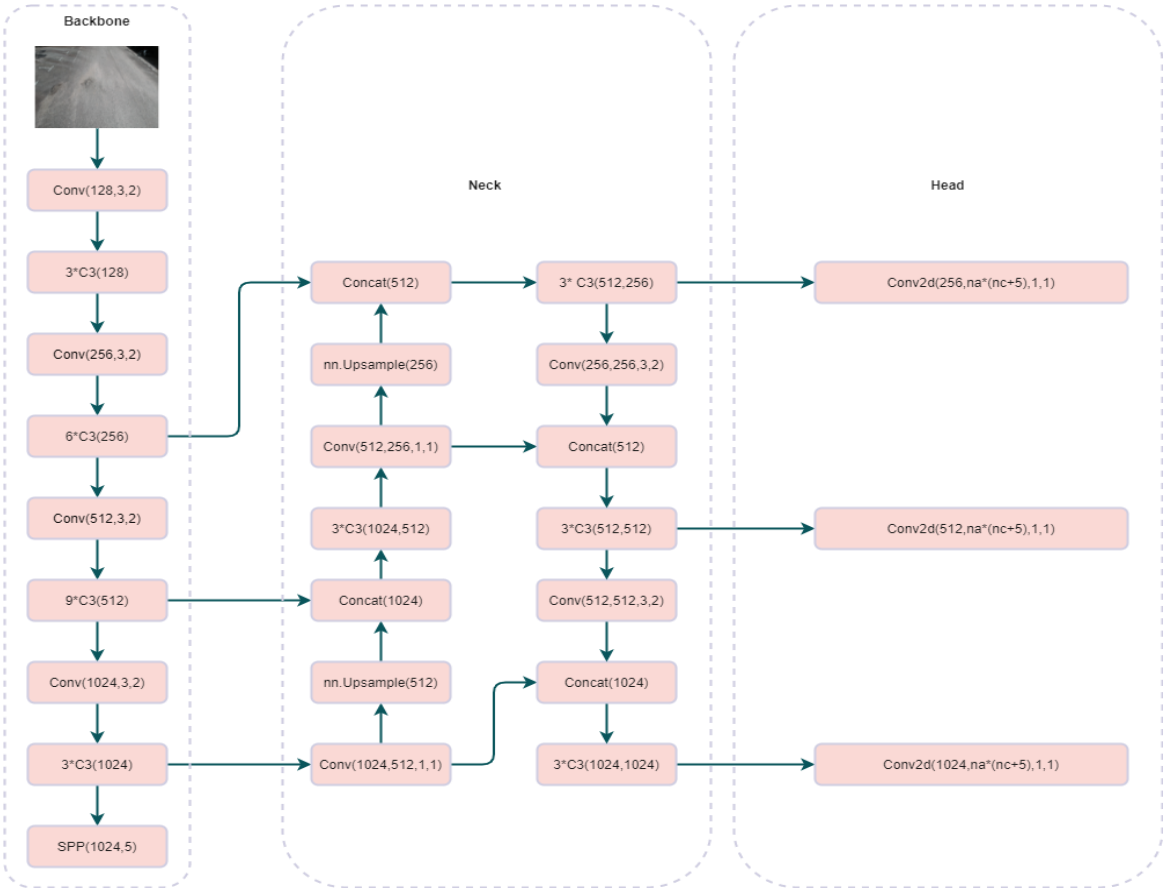

Figure 7. YOLOv5 architecture

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

The YOLOv5 architecture is meticulously crafted to surpass the efficiency of its predecessors, boasting an expanded and broader backbone network that adeptly captures an increased array of features from the input image. Augmenting its accuracy in detection, the architecture integrates a feature pyramid network alongside anchor boxes within the head network, empowering the model to excel in recognizing objects of diverse dimensions. Moreover, the strategic implementation of non-maximum suppression plays a pivotal role in purging superfluous detections, ultimately refining the final detection outcomes. Notably, as in Fig. 8, YOLOv5 comes in multiple iterations, including small, medium, large, and x-large versions, each tailored to cater to specific requirements and performance benchmarks.

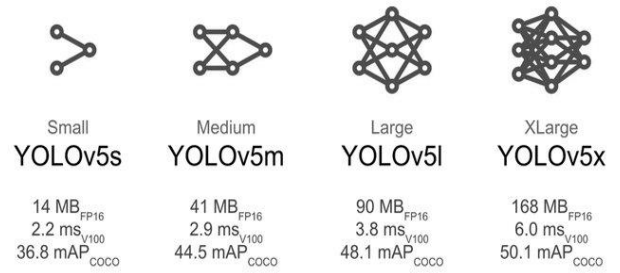

Figure 8. FP16 refers for the half floating-point precision, V100 is the inference time in milliseconds on the Nvidia V100 GPU, and mAP is based on the original COCO dataset in the YOLOv5 model sizes [14]

YOLOv5s: This particular variant stands as the most compact member within the family, encompassing approximately 7.2 million parameters. Its design is meticulously honed to cater to devices constrained by limited resources. It achieves this efficiency by incorporating a reduced count of convolutional layers and anchor boxes, distinguishing it from its counterparts in the same lineage.

YOLOv5m: With a parameter count of 21.2 million, this model assumes a medium-sized stance. It exhibits a higher number of convolutional layers and anchor boxes compared to YOLOv5s, effectively striking a commendable equilibrium between swiftness and precision.

YOLOv5l: Boasting an expansive parameter count of 46.5 million, this model takes the crown as the most capacious member within the YOLOv5 family. It notably encompasses an increased array of convolutional layers and anchor boxes compared to the YOLOv5 Medium variant.

YOLOv5x: Ranking as the most substantial among the quintet of models, this iteration commands the highest mean Average Precision (mAP). Despite its relatively longer processing time and a hefty parameter count of 86.7 million, it maintains a commendable swiftness that outpaces its counterparts.

YOLOv7

YOLOv7 architecture is influenced by its previous versions such as YOLOv4, scaled YOLOv4, and YOLO-R architecture. YOLOv7 (as in Fig. 9) provides several architectural enhancements that improve performance and accuracy. Similarly, the YOLOv7 architect all other YOLO family networks comprises of a Backbone, Neck, and Head. The YOLOv7's desired output is situated in the v7 architecture's head section. YOLOv7 beats all other known object detectors in terms of speed and accuracy in the 5 FPS to 160 FPS range, with the incredible accuracy of 56.8% AP of all real-time object detectors with 30 FPS or more [15].

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

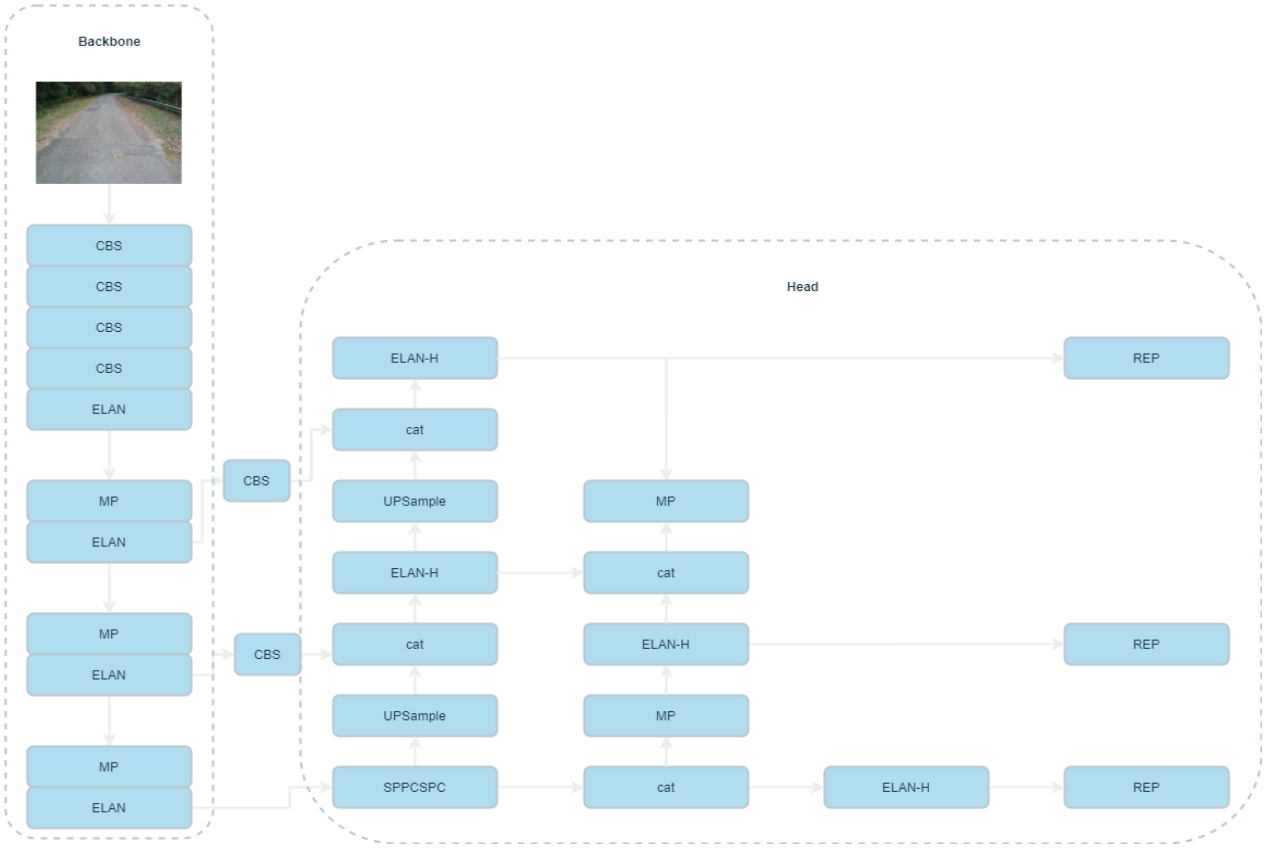

Figure 9. The structure of YOLOv7 model

Experiments, Results and Discussion:

YOLOv5 versions and YOLOv7 were implemented in this study due to the improved accuracy of this model compared to all of its predecessors. Both models produced faster inference speed, and these two models' robustness is a suitable fit for our experiments.

Different parameters are considered to evaluate the model's performance. To achieve reasonable accuracy, different experimental setup is being done. Various types of input sizes, number of epochs and fine-tuning the model's layers are some basic setups prepared before experiments of each model. During the first phase of the experiments, the input size was taken at 416x416, while for the 2nd experiment, the input size was taken at 640x640. The primary goal of varying input sizes in YOLO networks is to balance the trade-off between accuracy and speed. Larger input sizes usually result in higher accuracy but longer inference times, whereas smaller input sizes can result in shorter inference times but poorer accuracy.

Evaluation Matrices:

Precision indicates the ratio of accurate instances within all instances predicted as positive. Its computation incorporates TP (true positives) and FP (false positives) according to the formula: Precision = TP / (TP + FP).

Similarly, Recall reflects the percentage of relevant instances correctly identified by the model and is calculated using TP and FN (false negatives):

Recall = TP / (TP + FN).

The F1 score, which takes into account both Precision and Recall, is calculated by finding the harmonic mean of these measures:

F1 score = 2 * (Precision * Recall) / (Precision + Recall).

The mean Average Precision (mAP) is a widely used metric in object detection, which calculates the average of the Average Precision (AP) values for all classes. To compute the AP for each class, the area under the precision-recall curve is determined,

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

considering different thresholds. The precision and recall values are extracted from this curve. Eventually, the mAP is obtained by averaging these AP values across all classes.

Moreover, Giga Floating-point Operations Per Second (GFLOPs) serves as a metric to comprehend the model's capability in performing floating-point operations per second.

Evaluation Parameters:

The experiment is performed in Google Colab pro-environment with a memory size of 166GB and a total of 15GB of NVIDIA Tesla T4 GPU, and 15GB of RAM.

During the model training process, images are input at a size of 640x640 pixels, and Stochastic Gradient Descent (SGD) is employed as the optimization function. The training spans across 200 epochs, utilizing a batch size of 64 and an initially set default learning rate. The initial image input size adheres to the original YOLOv5 and YOLOv7 recommendations. A test run was conducted using an image pixel size of 416x416.

Comparative Study among YOLOv5 and YOLOv7 Versions:

For evaluating the models, based on the input size, each model of YOLOv5 is compared with the other two to find the best by considering all aspects. After that, all YOLOv5 models were compared with YOLOv7 models to find the most suitable one.

Table 3. Performance comparison among YOLOv5 different models

Models	Input Size	mAP (%)	Precision (%)	Recall(%)	F1 Score(%)	GFLOPS	Parameter
YOLOv5s	640	62.2	69.0	57.9	63.0	16.07	7.25 M
YOLOv5m	640	63.12	66.4	65.8	66.3	50.2	21.04 M
YOLOv5l	640	65.6	68.7	63.9	66.8	107.7	46.11 M
YOLOv5x	640	67.1	71.3	60.6	66.0	216.9	87.2 M

Table 4. Performance Comparison between YOLOv5s & YOLOv7

Models	Input Size	mAP (%)	Precision (%)	Recall (%)	F1 Score (%)	GFLOPS	Parameter
YOLOv5s	640	62.2	69.0	57.9	63.0	16.07	7.25 M
YOLOv7	640	70.4	75.4	62.4	68.3	105.1	37.2

Table 5. Performance Comparison between YOLOv5m & YOLOv7

Models	Input Size	mAP (%)	Precision (%)	Recall (%)	F1 Score (%)	GFLOPS	Parameter
YOLOv5m	640	63.12	66.4	65.8	66.3	50.2	21.04 M
YOLOv7	640	70.4	75.4	62.4	68.3	105.1	37.2

Table 6. Performance Comparison between YOLOv5l & YOLOv7

Models	Input Size	mAP (%)	Precision (%)	Recall (%)	F1 Score (%)	GFLOPS	Parameter
YOLOv5l	640	65.6	68.7	63.9	66.8	107.7	46.11 M
YOLOv7	640	70.4	75.4	62.4	68.3	105.1	37.2

Table 7. Performance Comparison between YOLOv5x & YOLOv7

Models	Input Size	mAP (%)	Precision (%)	Recall (%)	F1 Score (%)	GFLOPS	Parameter
YOLOv5x	640	67.1	71.3	60.6	66.0	216.9	87.2 M
YOLOv7	640	70.4	75.4	62.4	68.3	105.1	37.2

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Comparisons of Inference Time:

Table 8. Inference Comparison among YOLOv5 models

Models	Input Size	Inference Time(ms)	Pre-process Time(ms)	Per image (NMS)
YOLOv5s	640	12.1	0.3	1.0
YOLOv5m	640	14.1	0.3	1.0
YOLOv5l	640	27.6	0.5	1.0
YOLOv5x	640	29.6	0.6	1.3
YOLOv7	640	12.3	0.4	1.1

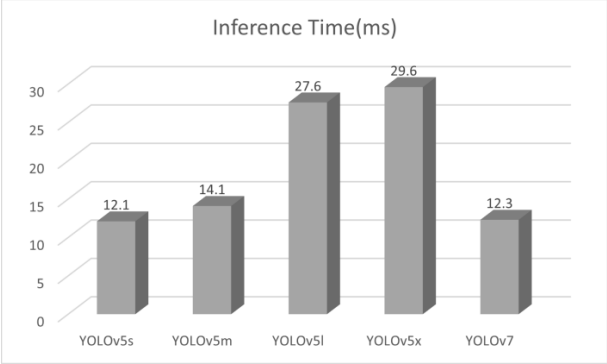

Figure 10. Inference time among YOLOv5 & YOLOv7 models as shown YOLOv5s took the lowest time than YOLOv5m, followed by YOLOv5l, YOLOv5x and YOLOv7, respectively.

The main reason behind the better performance in graph shown in Fig. 10, is the change of some significant parameters and approaches. Improving data annotation is also a considerable task in getting better accuracy. While cleaning the data, take care of the images with small bounding boxes or crowd, duplicate frames.

As shown in Figs. 7 and 9, our model's utilization of the PyTorch architecture enables a reduction in floating-point precision during training and inference. This shift is from 32 bits to 16 bits, resulting in a substantial acceleration of the overall process. Simultaneously, the inclusion of the CSP backbone and PA-Net neck, along with mosaic data augmentation and auto-learning bounding boxes, emerges as the most impactful factors contributing to the observed enhancement in performance.

Results and Discussion

Result Evaluation:

To understand the effectiveness of different YOLO versions, five experiments have been conducted based on different models proposed by the YOLO developers, using drone images to detect road cracks and potholes.

Each model has been tested with the same input size and learning rates. A few parameters are taken as standards for comparison among the models, such as Mean Average Precision (mAP), Recall, Precision, F1 score, and GFLOPs. While comparing YOLOv5s and YOLOv5m, the mAP has increased by 1.48%, while between YOLOv5m and YOLOv5l, the increased percentage is around 3.93, as well as the increase between YOLOv5l and YOLOv5x is about 2.29%. In the meantime, the Precision, Recall also fluctuates based on different models. Along with the mAP result, the F1 score has significant changes among different models; for example, there is a 5.24% increment between YOLOv5s and YOLOv5m, whereas the increase is about 0.75% between YOLOv5m and YOLOv5l. In the meantime, there is a reduction between YOLOv5l and YOLOv5x.

Moreover, while comparing different YOLOv5 models and the YOLOv7 model, it can be seen that YOLOv7 is 11.65%, 10.34%, 6.82% and 4.69% higher than the YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x respectively. As well as in F1 score of 2.25% on increased can be seen from YOLOv5l.

The images input size can significantly change the result of YOLO models, such as in our experiment YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

YOLOv7 shows mAP 59.5%, 61.5%, 61.9%, 63.8% and 67.2% respectably.

Furthermore, while inferencing the images, as mentioned in Table 8, the average inference time can be seen as a significant change among YOLOv5 models and YOLOv7, where the lowest inference time can be seen on YOLOv5s and the highest is in YOLOv5x, as expected due to the size of the network, YOLO5x is the maximum among others.

In our road damage detection experiment, the test results of YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and YOLOv7 indicated that all of these models performed remarkably well in recognizing and detecting road damages with remarkable confidence values as shown in the Fig. 11. The YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x versions detected all our selected road damage, such as cracks and potholes, with amazing precision. Furthermore, the YOLOv7 model offered more advanced features and enhanced performance in detecting road damage, making it a potential model for future applications. Overall, the excellent results obtained by all of the studied YOLO models demonstrate their effectiveness and potential in the field of road damage detection.

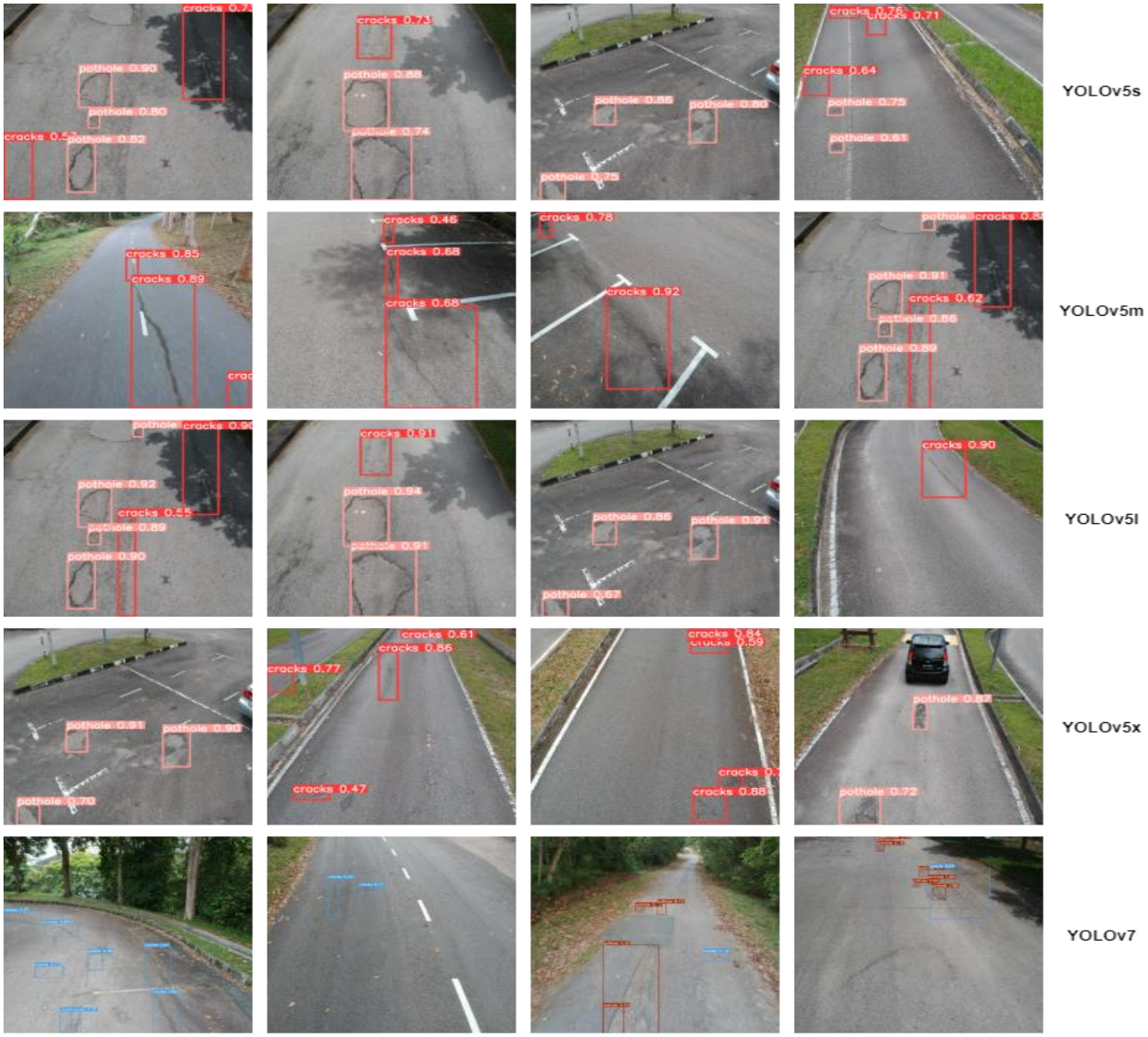

Figure 11. Inference YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x and YOLOv7

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Comparison against State-of-the-art Algorithms:

Table 9. Performance comparison with previous researchers

Models	Image size	F1 score
Faster R-CNN & Detectron2 [16]	512	51.4%
EfficientDet [17]	512	57.07%
Faster R-CNN with Resnet-50 [18]	320	53.6%
YOLOv5x [19]	416	57.10%
YOLOv5s- on our experiment	640	63.0%
YOLOv5m- on our experiment	640	66.3%
YOLOv5l- on our experiment	640	66.8%
YOLOv5x- on our experiment	640	66.0%
YOLOv7- on our experiment	640	68.3%

As in Table 9, after completing the experiments, the used YOLO models showed a significant improvement in overall performance in road crack and porthole detection on drone images. The used models (YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x and YOLOv7) showed F1 scores of 63%, 66.3%, 66.8%, 66% and 68.3%, respectively followed by the mAP values of 62.2%, 63.12%, 65.6%, 67.1% and 70.4%, which was about 15.5% improvement on F1 scores. The size of the input image had an extensive impact on the model's accuracy. When the size of the input image is increased, the model can capture more fine-grained details about the objects, resulting in improved object detection performance.

Conclusion

This study performed a competitive analysis of the latest and stable YOLO family networks' performance on new datasets of road cracks and potholes detection using drone images. This study's training and testing data are newly collected and noteworthy in road transport safety measurements. In our dataset, the YOLO network performed exceptionally well compared to the previous test done by different researchers. The confidence in detecting road cracks and potholes is remarkable. The various layers of YOLOv5 and YOLOv7 executed very well. YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, showed an acceptable result, while YOLOv7 functioned better than its previous versions due to the algorithmic functionality changes and the changes of hyperparameters in our architecture.

Road cracks and potholes are a significant concern regarding road safety. Our tested models improved the accuracy while keeping the parameters less. During the evaluation of the used models, it is notable that the input size and parameter numbers significantly change the model accuracy while the average inference time varies due to the size of the used models. Implementing YOLOv5 and YOLOv7 (based on the used dataset size) can show remarkable results in road safety management while keeping costs low and fastening inspection time. The integration of drone images has given a time and cost-effective approach for collecting data and identifying cracks in the images. The YOLO network has shown to be a worthwhile tool for identifying road cracks and potholes. The study has underlined the significance of precision road damage detection for assuring road safety and maintenance. The possibilities of using drone technology and YOLO networks on a grander scale for road inspection can significantly change road safety concerns. The study can be exceptionally beneficial for the company and organization involved in road inspection and maintenance. However, there is still an opportunity for improvement in detecting accuracy and speed.

Additional exploration could focus on refining the YOLO network to enhance its effectiveness across various scenarios. In general, this investigation contributes to the growing realm of studies involving the application of deep learning algorithms for the identification of road cracks and potholes. To enhance this work in the future, a two-step approach can be taken: initially identifying road defects and subsequently determining their potential harm through classification.

Authors' Declaration

- Conflicts of Interest: None.

- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and Images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.

- Ethical Clearance: The project was approved by the local ethical committee at University of Technology Malaysia.

Authors' Contribution Statement

This research was carried out in collaboration among the authors, A. H. S., generate the idea and was involved as a main contributor, while S. Z. M. H. contributed on analyzing ad proofreading along with H. and N.

References

1. Seo J, Duque L, Wacker J. Drone-enabled bridge inspection methodology and application. Autom Constr. 2018; 94: 112-126. https://doi.org/10.1016/j.autcon.2018.06.006
2. Ha J, Kim DS, Kim M. Assessing severity of road cracks using deep learning-based segmentation and detection. J Supercomput. 2022 May 22; 78(16): 17721–35. https://doi.org/10.1007/s11227-022-04560-x
3. Jang J, Yang Y, Smyth AW, Cavalcanti D, Kumar R. Framework of data acquisition and Integration for the detection of pavement distress via multiple vehicles. J Comput Civil Eng. 2017 Mar 1; 31(2): 04016061. https://doi.org/10.1061/(asce)cp.1943-5487.0000618
4. Syifa M, Park S, Lee CW. Detection of the Pine wilt disease tree candidates for drone remote sensing using artificial intelligence techniques. Eng. 2020 Aug 1; 6(8): 919–26. https://doi.org/10.1016/j.eng.2020.07.001
5. Mandirola M, Casarotti C, Peloso S, Lanese I, Brunesi E, Senaldi I. Use of UAS for damage inspection and assessment of bridge infrastructures. Int J Disaster Risk Reduct. 2022 Apr 1; 72: 102824. https://doi.org/10.1016/j.ijdrr.2022.102824
6. Awad JH, Majeed BD. Moving objects detection based on frequency domain. Baghdad Sci J. 2020 May 11; 17(2): 0556. https://doi.org/10.21123/bsj.2020.17.2.0556
7. Balakrishnan B, Chelliah R, Venkatesan M, Sah C. Comparative Study on Various Architectures of Yolo Models Used In Object Recognition. 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). 2022 Nov 4; https://doi.org/10.1109/icccis56430.2022.10037635
8. Yusro MM, Ali R, Hitam MS. Comparison of faster R-CNN and YOLOV5 for overlapping objects recognition. Baghdad Sci J. 2023; 20(3): 893-903. https://doi.org/10.21123/bsj.2022.7243
9. Wang Z, Zhu H, Jia X, Bao Y, Wang C. Surface Defect Detection with Modified Real-Time Detector YOLOv3. J Sens. 2022; 2022. https://doi.org/10.1155/2022/8668149
10. Redmon J, Santosh DHH, Ross G, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. arXiv (Cornell University). 2015 Jun 8. http://arxiv.org/abs/1506.02640
11. Norkobil Saydirasulovich S, Abdusalomov A, Jamil MK, Nasimov R, Kozhamzharova D, Cho YI. A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments. Sensors. 2023; 23(6): 3161. https://doi.org/10.3390/s23063161
12. Azurmendi I, Zulueta E, López-Guede JM, Azkarate J, González M. Cooktop sensing based on a YOLO object detection algorithm. Sensors. 2023 Mar 3; 23(5): 2780: https://doi.org/10.3390/s23052780
13. Dluznevskij D, Stefanovič P, Ramanauskaitė S. Investigation of YOLOV5 efficiency in iPhone supported systems. Balt J Mod Comput. 2021 Jan 1;9(3):07. https://doi.org/10.22364/bjmc.2021.9.3.07
15. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv (Cornell University). 2022 Jul 6; 9(3). http://arxiv.org/abs/2207.02696
16. Pham V, Pham C, Dang T. Road Damage Detection and Classification with Detectron2 and Faster R-CNN. Proceedings - 2020 IEEE Int Conf Big Data. Published online October 28, 2020: 5592-5601. https://doi.org/10.1109/BigData50022.2020.9378027
17. Tan M, Pang R, Le Q V. EfficientDet: Scalable and Efficient Object Detection. 2020 IEEE/CVF Conf Comp Vis Pattern Recognition (CVPR). Published online 2019: 10778-10787. https://doi.org/10.1109/CVPR42600.2020.01079
18. Vishwakarma R, Vennelakanti R. CNN Model & Tuning for Global Road Damage Detection. Proceedings - 2020 IEEE Int Conf Big Data. 2020. Published online March 17, 2021: 5609-5615. https://doi.org/10.1109/BigData50022.2020.9377902
19. Arya D, Maeda H, Ghosh SK, Toshniwal D, Omata H, Kashiyama T, et al. Global Road Damage Detection: State-of-the-art Solutions. Proceedings -

2024, 21(6): 2167-2181
https://doi.org/10.21123/bsj.2023.9027
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

2020 IEEE Int Conf Big Data. Published online November 17, 2020: 5533-5539. https://doi.org/10.1109/BigData50022.2020.9377790

YOLO: تحليل تنافسي لخوارزميات الكشف عن الأشياء الحديثة للكشف عن عيوب الطرق باستخدام صور الطائرات بدون طيار

أميت حسن سادهن[1]*، ستي زيتون محمد هاشم [1] ، حسين سماع [2] ، نورقيللا خميس [3]

[1]كلية الحاسبات، جامعة التكنولوجيا ماليزيا، جوهور، ماليزيا.
[2]مركز أبحاث مشترك بين سدايا وجامعة الملك فهد للذكاء الاصطناعي، جامعة الملك فهد للبترول والمعادن، المملكة العربية السعودية.
[3]كلية الهندسة الكهربائية، الجامعة التكنولوجية ماليزيا، جوهور، ماليزيا.

الخلاصة

يعد التحديد الفعال لعيوب الطرق مصدر قلق بالغ لما له من اثر على السلامة على الطرق وصيانة البنية التحتية. يستخدم هذا البحث الصور الملتقطة بطائرات بدون طيار وخوارزميات متقدمة للكشف عن الأشياء لتسريع عملية التعرف على العيوب، مع التركيز بشكل خاص على تحديد الخوارزمية المثالية للكشف السريع والدقيق. تظل أهمية اكتشاف عيوب الطريق في الوقت المناسب، وهو أمر بالغ الأهمية للتخفيف من المخاطر المحتملة، أمرًا أساسيًا. تحليل مقارن شامل لخوارزميات الكشف عن الكائنات المعاصرة، بما في ذلك YOLOv5s وYOLOv5m وYOLOv5l وYOLOv5x و YOLOv7. تسلط نتائج هذه الدراسة الضوء على YOLOv7 باعتباره الأكثر كفاءة، مع mAP ملحوظ بنسبة 68.3%، يليه (66.8%) YOLOv5l، و(66.3%) YOLOv5m، و YOLOv5x (66%)، و YOLOv5s (63%). يؤدي دمج الصور المشتقة من الطائرات بدون طيار، والتقاط التدرجات المميزة، إلى تعزيز دقة اكتشاف العيوب بشكل كبير. وبعيدًا عن السلامة على الطرق، تقدم هذه الدراسة رؤى قيمة لممارسي الرؤية الحاسوبية والتعلم الآلي. ومن خلال ربط الابتكار التكنولوجي بالتنفيذ العملي، فإنه يحمل القدرة على تعزيز السلامة على الطرق وجودة البنية التحتية للنقل واستخدام تكنولوجيا الطائرات بدون طيار الثورية.

الكلمات المفتاحية: الشبكة العصبية التلافيفية، YOLO، صور الطائرات بدون طيار، CSPDarknet، حجم الإدخال.