Road Surface Anomaly Detection Using YOLOv4 Tiny Model with Potential Embedded Framework

Mst Ayesha Khatun, Md Facklasur Rahaman, Dong-Seong Kim, Jae Min Lee and Jung-Hyeon Kim
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
(20226138, facklasur,dskim, ljmpaul, trizkim)@kumoh.ac.kr

***Abstract*—In the realm of computer vision, real-time object detection plays a pivotal role in a plethora of applications, from autonomous vehicles to smart surveillance systems. This paper presents an in-depth exploration of utilizing the YOLOv4-tiny model for the task of object detection, specifically focusing on road surface imagery classification into five distinct classes namely plain, crack, pothole, black ice, and obstacle. The focus on road surface imagery introduces challenges related to varying roads and weather conditions making the anomaly detection both important and challenging. By leveraging YOLOv4-tiny's lightweight architecture, we demonstrate its effectiveness in achieving road surface anomaly detection accuracy on par with larger models. On top of the great detection accuracy already achieved, we propose an enhancement of the existing model that has the potential to gain commendable compactness, making it an ideal candidate for deployment on edge devices, including resource-constrained platforms such as mobile phones. We limit the results presented in this paper to the great detection results achieved for the mentioned five classes. We outline the framework of the potential enhancement of our model towards being embedded in mobile devices end of the paper.**

***Index Terms*—Object Detection, YOLOv4-tiny, Embedded Systems, Lightweight Models**

I. Introduction

YOLOv4, the fourth iteration of the You Only Look Once (YOLO) series developed by Alexey Bochkovskiy, represents a significant advancement in real-time object detection [1] [2] [3]. Released in 2020, YOLOv4 builds upon the strengths of its predecessors, incorporating several innovations to enhance both accuracy and speed. YOLOv4's architecture is a fusion of various techniques, including the CSPDarknet53 backbone [4], PANet, PANet Neck [5], and SAM block [6], which collectively enable better feature extraction and multi-scale detection. Darknet is a framework that includes a Convolutional Neural Network (CNN) [7] [8] architecture for various computer vision tasks, particularly object detection. The introduction of the CIOU loss function [9] in YOLOv4 further refines the model's bounding box predictions. By addressing challenges posed by previous versions, YOLOv4 delivers state-of-the-art performance in object detection tasks, outperforming its predecessors and competing models.

The YOLOv4 Tiny model marks a significant milestone in the evolution of real-time object detection algorithms. Introduced as a compact and efficient variant of the YOLOv4 architecture, YOLOv4 Tiny addresses the demand for object detection models that can be seamlessly integrated into edge devices with limited computational resources. This model achieves a remarkable balance between size and performance, enabling its deployment on edge devices without sacrificing detection accuracy. By adopting a streamlined architecture that maintains the core principles of YOLO—such as anchor-based detection and multi-scale feature extraction—YOLOv4 Tiny optimizes the model for resource-constrained environments. Its reduced size and computational requirements make it ideal for applications such as surveillance cameras, drones, robotics, and IoT devices, where real-time object detection is essential.

YOLO serial methods are generally complicated network structures because of larger number of network parameters. They have limited computing power and limited memory, and require real-time object detection for some mobile devices and embedded devices in everyday applications. The available computing resources to do such challenging tasks are limited to a combination of low-power embedded GPUs or even just embedded CPUS with limited memory. Therefore, it is very difficult task to do the real-time object detection on embedded devices and mobile devices. In order to solve this problem, the lightweight methods have come in, that have comparatively simpler network structure and fewer parameters. Hence the required computing resources and memory are at managable limits and they have faster detection speed. Because of their smaller size, they are more suitable for deploying on mobile devices and embedded devices. The detection accuracy is tempered because of all these imposed restrictions but still meets the demand. Their widespread applications include vehicle detection, pedestrian detection, bus passenger object detection, agricultural detection, human abnormal behavior detection, etc.

In this paper we propose an innovative object detection model based on the YOLOv4 tiny methodology, designed to effectively discern five distinct road surface classes namely plain, crack, pothole, black ice, and obstacle. Our proposed model not only demonstrates great accuracy but also possesses the potential for real-time object detection because of its smaller size suitable to be deployed in mobile devices. We presented the object detection result using YOLOv4 algorithm for the mentioned five classes and proposed a framework for further enhancing the efficacy of the model by embedding it to mobile devices.

The rest of the paper is organized as follows: section 2 has the details of the base model of our framework i.e. YOLOv4 model, section 3 gives the elaborated system model that we proposed, section 4 has our object detection results for the mentioned five classes and the relevant discussion, section 5 has the details of our future work that we are currently working on building on top of the object detection results that we got and the last and final section of the paper has the conclusion followed by references.

II. YOLOv4 Network Architecture

Yolov4-tiny method is designed based on Yolov4 that has a speed of object detection about 371 Frames per second using 1080Ti GPU with the accuracy that meets the demand of the real application. Unlike Yolov4 method, the Yolov4-tiny method uses CSPDarknet53-tiny network as backbone network instead of the CSPDarknet53. The CSPDarknet53-tiny network uses the CSPBlock module in cross stage partial network instead of the ResBlock module in residual network. This makes it possible that the gradients propagate in two different network paths to increase the correlation difference of gradient information. The CSPBlock module enhances the learning ability of convolution network $(10-20)\%$ as compared to ResBlock module. CSPBlock module removes the computational bottlenecks that have higher amount of calculation and thus improves the accuracy of Yolov4-tiny method in the case of constant or even reduced computation. To simplify the computation process, YOLOv4-tiny employs the LeakyReLU activation function within the CSPDarknet53-tiny network, foregoing the Mish activation function used in YOLOv4. Unlike YOLOv4, it omits spatial pyramid pooling and path aggregation. YOLOv4-tiny employs two scale-specific feature maps (13x13 and 26x26) to predict detections.

To address the issue of redundant bounding boxes during prediction, a confidence threshold is introduced. If a bounding box's confidence score exceeds this threshold, the box is retained; otherwise, it's discarded. The confidence score is determined by comparing predicted and ground truth bounding boxes through the Intersection over Union (IoU) measure. The objectness score reflects the proximity of the predicted box to the ground truth. YOLOv4-tiny employs the same loss function as YOLOv4, encompassing confidence, classification, and bounding box regression losses. The confidence loss function differentiates between bounding boxes responsible for detection (with corresponding weights). The classification loss involves comparing predicted and actual class probabilities. Weighted by parameters, these loss components contribute to the overall optimization. This approach streamlines bounding box predictions and class assignments, enhancing the efficiency of YOLOv4-tiny's object detection framework.

The loss function calculates the difference between predicted and ground truth bounding boxes' dimensions, widths, and heights. This difference is scaled by factors such as the intersection over the union between predicted and truth bounding boxes, the diagonal distance of the boxes, and the truth width and height.

III. Our Proposed Object Detection Model

We gathered a diverse dataset comprising images of road surfaces captured under five different safety and weather conditions namely plain, crack, pothole, black ice and obstacle. The dataset was manually annotated with bounding boxes and class labels for the five targeted classes. The YOLOv4 Tiny architecture was selected due to its compact size and ability to maintain a balance between accuracy and speed. It consists of a backbone network, feature extraction layers, and detection layers. The dataset was split into training and validation sets. Our system model is depicted in Figure 1.

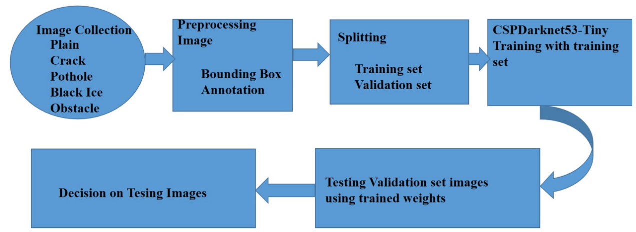

Fig. 1. System model

The system model encompasses a streamlined process for effective road surface detection utilizing the YOLOv4 Tiny neural network. Commencing with image collection, diverse road surface images depicting plain, crack, pothole, black ice, and obstacle scenarios are gathered. These images undergo annotation and bounding box labeling during preprocessing, essential for training the YOLOv4 Tiny model. The dataset is then divided into training and validation sets, facilitating model training and performance evaluation. The Darknet neural network is employed to train on the training dataset, enabling it to identify and localize different road surface elements. The trained model is tested on validation images with the highest percentage value in favor of a class being the final decision rule.

The model was trained from the scratch using annotated images i.e. we did not utilize transfer learning by initializing the model with pre-trained weights. Darknet provides the foundation for training and deploying the YOLOv4 tiny model. The framework supports various neural network architectures and optimization techniques and is well-suited for training custom YOLO models on specific datasets. The training process involved adjusting hyperparameters and optimizing the model to achieve accurate predictions. The model's performance was assessed in terms of percentage accuracy with a focus on real-time application feasibility.

IV. Result and Discussion

The performance of the YOLOv4 Tiny model in object detection for different road surface classes was evaluated, and the average accuracy of the model's decisions for each class was analyzed. The obtained results are presented in Table 1, which displays the accuracy percentages for each class.

The YOLOv4 Tiny model demonstrated high accuracy in detecting various road surface classes. The class "Plain" achieved the highest accuracy of 96%, indicating the model's proficiency in recognizing normal road surfaces. "Crack" and "Pothole" classes also exhibited commendable accuracy levels

TABLE I
AVERAGE OBJECT DETECTION ACCURACY OF OUR PROPOSED MODEL FOR DIFFERENT CLASSES

Class	Average Accuracy
Plain	96%
Crack	93%
Pothole	93%
Black Ice	82%
Obstacle	94%

of 93%, showcasing the model's ability to identify common road surface anomalies. The challenge of collecting sufficient black ice images was encountered during the training process. The "Black Ice" class achieved an accuracy of 82%, which is relatively lower than other classes. This decrease in accuracy can be attributed to the scarcity of black ice images in the training dataset. Obtaining authentic images of black ice is inherently challenging due to its transient and often dangerous nature. As a result, the limited availability of diverse black ice images hinders the model's exposure to this critical road surface condition, impacting its overall accuracy. Lastly, the "Obstacle" class achieved an accuracy of 94%, highlighting the model's success in detecting obstructions on the road.

The obtained results affirm the YOLOv4 Tiny model's effectiveness in road surface detection. It demonstrates promising accuracy levels across diverse road surface conditions.

V. WORK IN PROGRESS TOWARDS EMBEDDING THE MODEL INTO MOBILE DEVICES

In our forthcoming endeavors, we are focused on advancing the deployment capabilities of our specialized object detection model tailored for road surface analysis. The future roadmap of our project is outlined in the block diagram given by Figure 2.

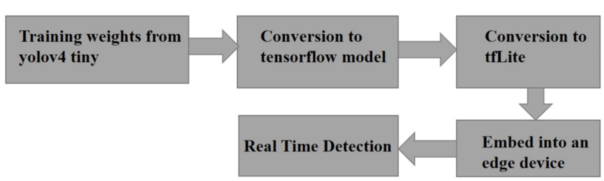

Fig. 2. work in progress roadmap towards an embedded object detection model

Our primary objective is to transition from the existing YOLOv4 model, fine-tuned for detecting five distinct classes of road conditions—plain, crack, pothole, black ice, and obstacle—to a TensorFlow-compatible variant. This conversion process involves adapting the model's architecture and transferring the meticulously learned weights, ensuring the retention of valuable insights derived from training. Subsequently, we plan to harness the capabilities of TensorFlow Lite (tfLite), a specialized framework designed for efficient deployment on resource-constrained devices like mobile phones. By converting the TensorFlow model to a tfLite model, we will achieve a streamlined version optimized for real-time road surface analysis while upholding a high level of accuracy. This compact and efficient tfLite model will be well-suited for integration into mobile devices, facilitating on-the-go road condition detection. By adopting this approach, we anticipate empowering mobile devices to swiftly and accurately identify specific road surface types—plain, crack, pothole, black ice, and obstacle—in diverse environmental conditions, consequently fostering safer and more efficient transportation systems.

VI. CONCLUSION

In conclusion, our object detection model, based upon the YOLOv4 tiny architecture, has exhibited commendable performance in accurately identifying the distinct road surface anomalies belonging to five distinct classes namely plain, crack, pothole, black ice, and obstacle. The only disparity in performance was the black ice image detection that showed slightly lower detection accuracy because of the scarcity of training images. We have already set our roadmap for converting our current model to a tfLite model to be able to embed the model into a mobile device. Our intermediate results showing great signs in favor of achieving our target. Once we achieve that the final model will be able to do real time detection of road surface anomalies.

VII. ACKNOWLEDGMENT

This research was supported by Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003), and by MSIT under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) supervised by the IITP, and the project for Industry-University-Research Institute platform cooperation RD funded Korea Ministry of SMEs and Startups in 2022.(S3310783).

REFERENCES

[1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[2] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved yolov4-tiny," *arXiv preprint arXiv:2011.04244*, 2020.

[3] A. M. Roy, R. Bose, and J. Bhaduri, "A fast accurate fine-grain object detection model based on yolov4 deep neural network," *Neural Computing and Applications*, pp. 1–27, 2022.

[4] M. Mahasin and I. A. Dewi, "Comparison of cspdarknet53, cspresnext-50, and efficientnet-b0 backbones on yolo v4 as object detector," *International Journal of Engineering, Science and Information Technology*, vol. 2, no. 3, pp. 64–72, 2022.

[5] F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Single-stage uav detection and classification with yolov5: Mosaic data augmentation and panet," in *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2021, pp. 1–8.

[6] X. Ma, K. Ji, B. Xiong, L. Zhang, S. Feng, and G. Kuang, "Light-yolov4: An edge-device oriented target detection method for remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10 808–10 820, 2021.

[7] R. Akter, V.-S. Doan, J.-M. Lee, and D.-S. Kim, "Cnn-ssdi: Convolution neural network inspired surveillance system for uavs detection and identification," *Computer Networks*, vol. 201, p. 108519, 2021.

[8] A. J. Garcia, A. Aouto, J.-M. Lee, and D.-S. Kim, "Cnn-32dc: An improved radar-based drone recognition system based on convolutional neural network," *ICT Express*, vol. 8, no. 4, pp. 606–610, 2022.

[9] C. Guo, X.-l. Lv, Y. Zhang, and M.-l. Zhang, "Improved yolov4-tiny network for real-time electronic component detection," *Scientific reports*, vol. 11, no. 1, p. 22744, 2021.