Mathematical Biosciences and Engineering

MBE, 20(1): 283–295.
DOI: 10.3934/mbe.2023013
Received: 26 July 2022
Revised: 30 August 2022
Accepted: 09 September 2022
Published: 30 September 2022

http://www.aimspress.com/journal/MBE

Research article

Effect of Gaussian filtered images on Mask RCNN in detection and segmentation of potholes in smart cities

Auwalu Saleh Mubarak[1,2,*], Zubaida Said Ameen[1,3] and Fadi Al-Turjman[1,4]

[1] Artificial Intelligence Engineering Dept., AI and Robotics Institute, Near East University, Mersin 10, Turkey
[2] Electrical Engineering Dept., Kano University of Science and Technology, Wudil, Kano, Nigeria
[3] Biochemistry Dept., Maitama Sule University, Kano, Nigeria
[4] Research Center for AI and IoT, Faculty of Engineering, University of Kyrenia, Kyrenia, Mersin 10, Turkey

* **Correspondence:** Email: auwalusaleh.mubarak@neu.edu.tr.

Abstract: Accidents have contributed a lot to the loss of lives of motorists and serious damage to vehicles around the globe. Potholes are the major cause of these accidents. It is very important to build a model that will help in recognizing these potholes on vehicles. Several object detection models based on deep learning and computer vision were developed to detect these potholes. It is very important to develop a lightweight model with high accuracy and detection speed. In this study, we employed a Mask RCNN model with ResNet-50 and MobileNetv1 as the backbone to improve detection, and also compared the performance of the proposed Mask RCNN based on original training images and the images that were filtered using a Gaussian smoothing filter. It was observed that the ResNet trained on Gaussian filtered images outperformed all the employed models.

Keywords: Mask RCNN; pothole; computer vision; smart cities; object detection; Gaussian filter

1. Introduction

Several faults such as potholes, cracks and uneven surfaces on roads cause a significant amount of loss of lives in the world. The roads as backbones of transformations contribute to the nation's economic growth. Since 2011, drivers have spent more than $3 billion on car repairs due to pothole damage [1]. From the fiscal year 2013 through the fiscal year 2021, the yearly number of potholes fixed in San Antonio, Texas was over 100,520, with COVID-19 accounting for 80,937 of those. This is a significant toll, and the paper suggests that many cases go undetected. The United States' road network is extensive: there are roughly 746,100 miles of road in the country [2]. Deep learning has produced several advances and successes in complicated feature extraction and target recognition in recent years. Simultaneously, the combination of deep learning models and computer vision has advanced Artificial Intelligence (AI) in recent years [3–9].

To improve the single-stage object detector's speed and accuracy, YOLOV3 was introduced, which is a combination of improved ResNet as backbone and object box prediction [10]. Convolutional neural networks (CNNs) are used to teach computers how to detect and segment objects with a high complexity but little contour and edge information. Maskrcnn is a good model for industrial image identification because it has a clear network topology, high resilience, and simultaneous recognition and segmentation processing [11].

Techniques such as sensor-based [12,13], image processing and machine learning-based [14–21], laser-based 3D reconstruction and stereo vision-based [22–28], are some of the approaches used to automate the pothole detection process on roads. Sensor-based systems employ vibration sensors to detect potholes. The accuracy of detecting potholes may be impacted by false-positive and false-negative readings from the vibration sensor interpreting joints on roads as potholes or not detecting potholes in the centre of a lane, respectively.

Finding an efficient model that will detect and segment potholes with high accuracy and speed is very important, several studies were carried out to perform pothole detection but are based on classification models, Fast RCNN, Faster RCNN and so on. In this study, the following contributions were made to improve pothole detection and segmentation at the same time with higher accuracy and speed of detection.

1) Effect of Gaussian Smoothing Filter on Mask RCNN with different backbone was studied.

2) New Mask RCNN with ResNet-50 and MobileNetv1 backbone was proposed.

3) Gaussian smoothing filter was employed to improve the training images as well as the model's performance.

4) The performance of the models was compared using mean average precision and recall to find the best performing model.

5) The ResNet-50 trained using the Gaussian filtered images outperformed the remaining models employed in the study.

2. Related works

With the current availability of low-cost cameras and improved image processing techniques, interest have grown in developing new pothole detection models. Traditional image processing algorithms are accurate, but they need time-consuming tasks like manually extracting features and changing image processing parameters. To restore the pavement surface, 3D reconstruction methods acquire 3D road data. A trained model for detecting potholes in 2D digital pictures was created using machine learning techniques. In order to enhance the accuracy of ML techniques, experts must manually extract attributes. Deep convolutional neural network operations are used in DL techniques to automate data extraction and categorization in real time [29,30].

DeepMask which produces a mask on the target object instance segmentation found to have low boundary segmentation accuracy is recommended by [31]. The first end-to-end instance segmentation framework proposed by [32], Full Convolutional Instance Segmentation (FCIS) improve the position-sensitive score map by predicting both the bounding box and instance segmentation, the drawback of this model was, that it was not able to predict the boundaries of occluded objects efficiently [33]. Mask RCNN framework, which is an algorithm with relatively improved instance segmentation results among existing segmentation algorithms was proposed [7,34]. A mask R-CNN outperforms a faster R-CNN in terms of performance [7]. Cucumbers with a similar colored body and leaves are properly detected by a mask R-CNN [35]. It also distinguishes between six distinct types of culinary equipment and photovoltaic plant borderlines [36,37]. A mask R-CNN has a higher average precision (AP), allowing it to quickly evaluate accident indemnities by detecting the level of vehicle damage [38]. It has an AP of 98% for sorting various-sized hardwood planks [39], hence it is utilized to categorize the planks for manufacturers development of enhanced image processing methods and the availability of cheap camera sensors have fueled the development of model-based pothole detection systems [40]. Traditional image processing algorithms are accurate, but they need time-consuming tasks like manually extracting features and changing image processing parameters. To restore the pavement surface, 3D reconstruction methods acquire 3D road data. A trained model for detecting potholes in 2D digital pictures was created using machine learning techniques. In order to enhance the accuracy of ML techniques, experts must manually extract attributes. Deep convolutional neural network operations are used in DL techniques to automate data extraction and categorization in real time [29].

3. Pothole detection

3.1. Mask RCNN

Mask RCNN, an instance segmentation algorithm model that can segment objects at the pixel level while identifying targets, was used in this work to detect potholes. Faster RCNN, Region of Interest alignment algorithm (ROIAlign) and Feature Pyramid Networks (FPN) made up the Mask RCNN topology [41]. The Mask R-CNN was not meant to coordinate network inputs and outputs pixel-by-pixel. The way RoIPool, the defacto basic process for attending to instances, conducts coarse spatial quantization for feature extraction exemplifies this. The network structure is depicted schematically in Figure 1.

The FPN and RPN of the backbone network execute multi-dimensional feature extraction and information fusion, while the RPN also produces and provides target candidate areas based on extracted feature maps and classifications. Finally, the ROIAlign is used to correct the target region, which is then integrated with FCN to perform target instance segmentation [42]. RoIAlign is a straightforward, quantization-free layer that accurately stores precise spatial positions. Under more stringent localization measures, it increases mask accuracy by 10 to 50%. To detect the category, we decouple mask and class prediction and depend on the network's RoI classification branch. Mask RCNN outperforms all state-of-the-art single-model results on the COCO instance segmentation challenge based on ablation experiments conducted by [7].

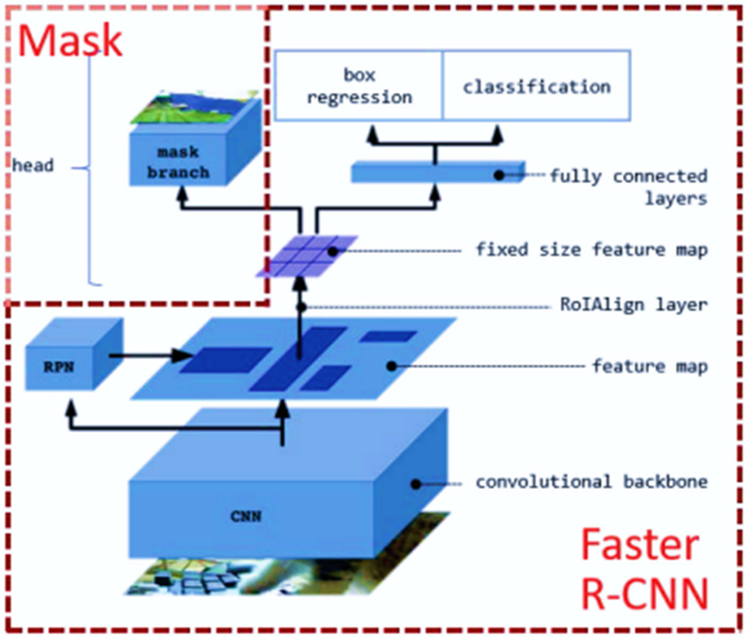

Figure 1. Mask RCNN structure [43].

3.2. MobileNet

The MobileNet [44] model is a network model in which the basic unit is depthwise separable convolution. It contains two levels: depthwise and point convolutions, which are regarded separately as convolution layers. Each dense block layer's input feature maps are a superposition of the preceding convolution layer's output feature maps. Between two dense blocks in DenseNet, there is a transition layer. The transition layer uses a 1×1 convolution kernel to decrease the number of input feature maps. MobileNet does not have a transition layer and instead relies on a convolution layer rather than a pooling layer. To minimize the size of the feature map, the convolution layer directly convolutes the output feature map of the preceding point convolution layer with stride 2. The architecture is presented in Figure 2.

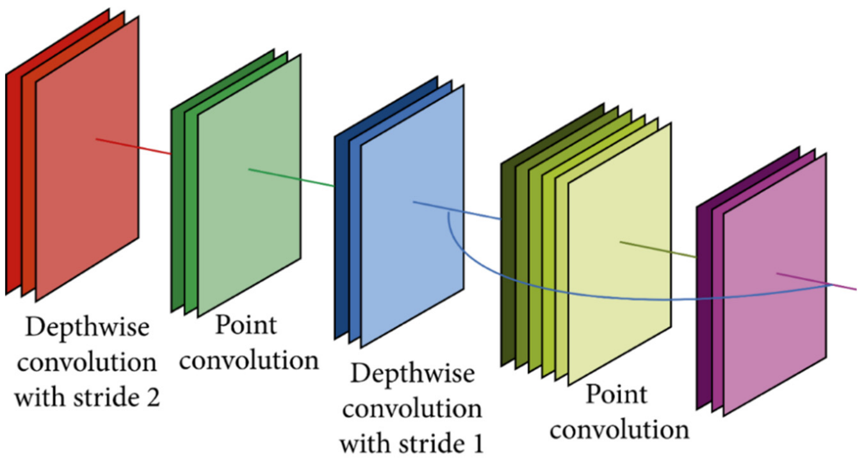

Figure 2. MobileNet architecture [45].

3.3. ResNet

Machine learning experts add extra layers while working with deep convolutional neural networks to tackle an issue in computer vision. These extra layers aid in the faster resolution of complicated issues since separate layers may be taught for different tasks to produce highly accurate outcomes. While the number of stacked layers might enhance the model's characteristics, a deeper network can reveal the degradation issue. Overfitting has not caused this deterioration. It might be caused by the network's setup, optimization algorithm, or, more crucially, the issue of vanishing or exploding gradients. To increase the accuracy of the models, deep residual nets incorporate residual blocks. They solve the problem of disappearing gradients by creating an alternate path for the gradient to follow. They also allow the model to learn an identity function, which assures that the model's upper layers perform equally well as the lower levels. ResNet was built specifically to address this issue [8,9]. The architecture is presented in Figure 3.

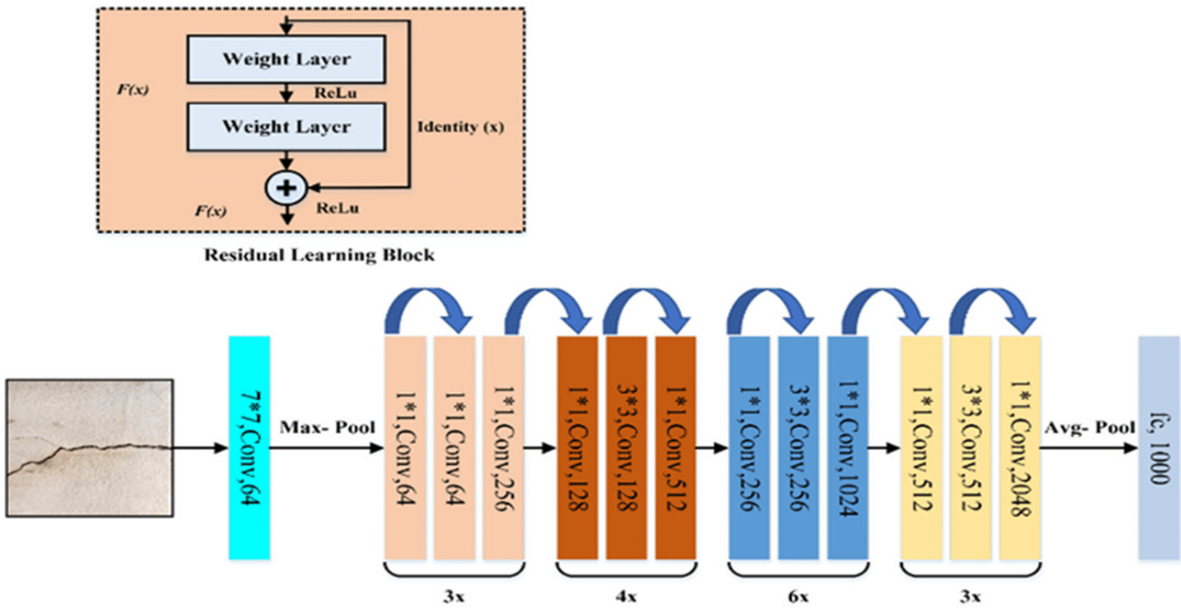

Figure 3. ResNet-50 architecture [46].

The ResNet-50 design is based on the Resnet34 model, however, each building block is made up of a stack of three layers rather than two. This model generates 3.8 billion FLOPS and is substantially more accurate than the 34-layer ResNet model. A 50-layer design was created by replacing each of the previous 2-layer blocks with a 3-layer bottleneck block [46].

3.4. Gaussian filtering (smoothing)

Gaussian filters are a class of linear smoothing filters with weights chosen according to the form of the Gaussian function [47]. Eq (1) is a very good filter for eliminating noise taken from a normal distribution the Gaussian smoothing filter.

$$G(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

Determines the width of the Gaussian. The two-dimensional discrete Gaussian zero mean function, Eq (2) is used as a smoothing filter for image processing.

$$g[i,j] = e^{-\frac{(i^2+j^2)}{2\sigma^2}} \quad (2)$$

3.5. Dataset and data preprocessing

665 pothole images from [48] were used in this study, image annotation was carried out using VGG Image Annotator (VIA) [49], and polygons were drawn around the potholes on each image. Before the annotation, the images were made into sets, a Gaussian smoothing filter was applied to one set of the images. 80% of the images were used for training, and 20% for validation. Samples of smoothening filtered and original potholes are presented in Figure 4.

Figure 4. Original and smoothen images.

3.6. Model training

In this study, Mask-RCNN with two different backbones ResNet-50 and MobileNetv1 was employed to detect and segment potholes and to compare the performance of the proposed detection model, weight of previously trained models was adopted using the transfer learning technique, The models were trained for two different scenarios, in the first scenario the model was trained on the original images using ResNet-50 and MobileNetv1 as the backbone, while in the second scenario, the model was trained on images which were smoothen using Gaussian smoothing filter with above-mentioned backbones. Before training the model, the data was augmented to increase the number of images and also the model performance. The data augmentation will also prevent the model from overfitting. The models' hyperparameters were tuned to learning rate = 0.001, learning-momentum = 0.9, weight-decay = 0.0001, detection min confidence = 0.9, steps pre epoch num_classes = 100, maskpool_size pool_size = 14, validation_steps = 100, epoch = 50. The training was carried out using a system with GeForce Nvidia GTX1080, RAM 16 Gb and processor i7. Anaconda environment with jupyter notebook tensorflow = 1.4.0 and keras = 2.0.8 was used for the training environment. The performance evaluation used for the models is mean Average Precision (mAP) and Recall at a threshold of 50 and 75% detection levels. The framework for the employed pothole detection and segmentation is presented in Figure 5.

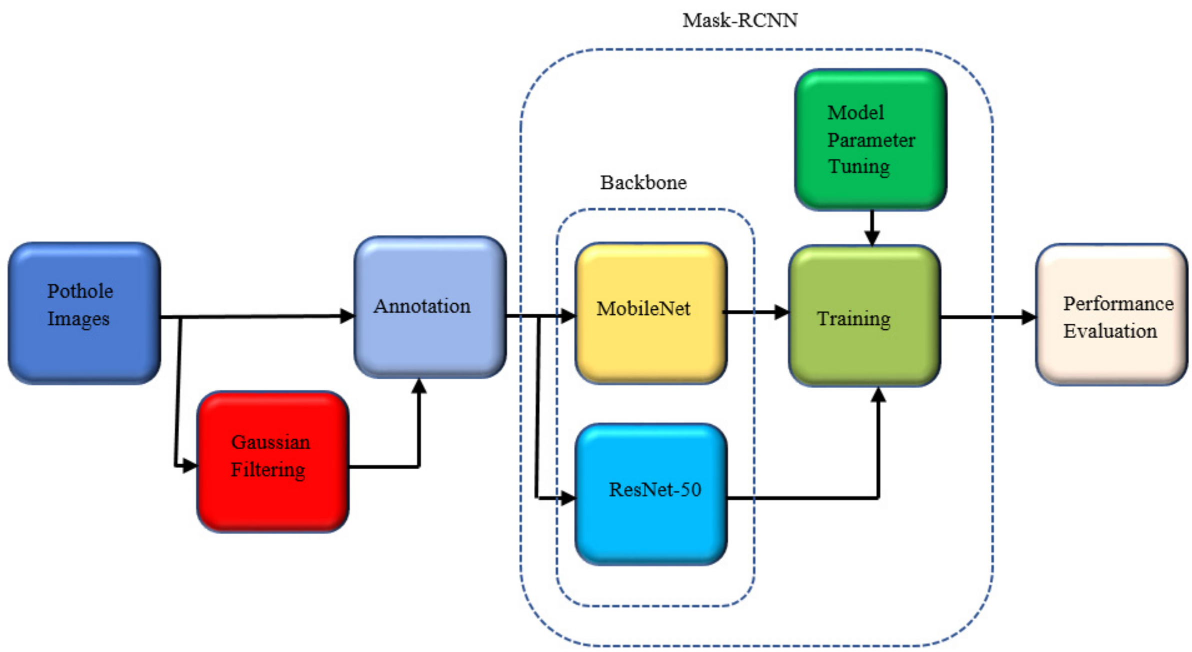

Figure 5. Framework for pothole detection.

4. Results and discussion

In this study, Mask RCNN was employed with different backbone networks and enhanced images using a Gaussian filter to detect and segment potholes, from the study, models' performance based on AP and Recall were compared to find the best performing model. In Table 1, it was observed that the best performing model which is ResNet-50 with Gaussian filtered images achieves the highest performance at 94.43, 96.26 and 98.10% with respect to mAP, mAP-75 and mAP-50, the results also show 2% increase in performance compared to the ResNet-50 trained with original images. The least performing model is MobileNet with 59.38, 70.30 and 75.75% mAP, mAP-75 and mAP-50 respectively. The highest recall was achieved by the ResNet-50 trained using Gaussian Filtered images as presented in Table 2. The results are presented graphically in Figures 6 and 7.

Table 1. Performance of the employed models based on average precision.

Backbone	mAP (%)	mAP-75 (%)	mAp-50 (%)
MobileNet	59.38	70.30	75.75
ResNet-50	91.22	94.73	95.63
Gaussian Filter + MobileNet	61.66	71.56	77.56
Gaussian Filter + ResNet-50	94.43	96.26	98.10

Table 2. Performance of the employed models based on Recall.

Backbone	Recall (%)	Recall-75 (%)	Recall-50 (%)
MobileNet	53.96	68.12	70.97
ResNet-50	88.19	91.51	93.60
Gaussian Filter + MobileNet	56.69	70.12	71.97
Gaussian Filter + ResNet-50	90.19	94.15	95.36

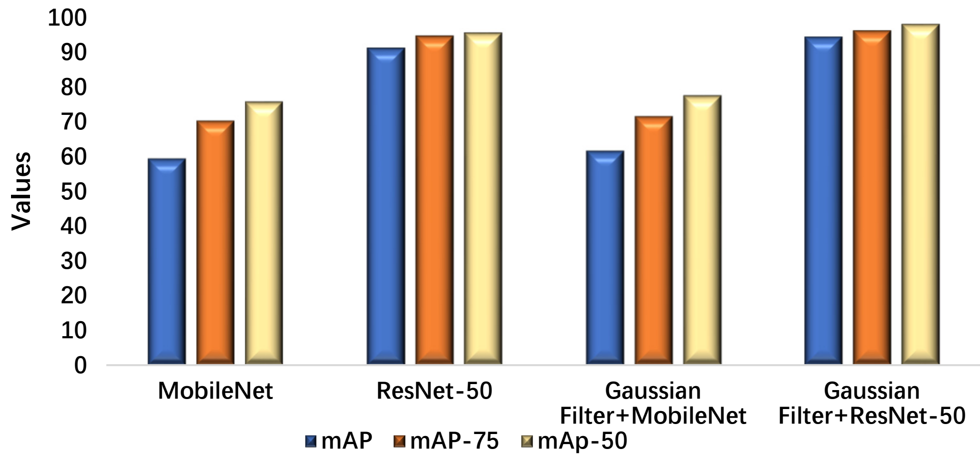

Figure 6. Performance of the employed models based on average precision.

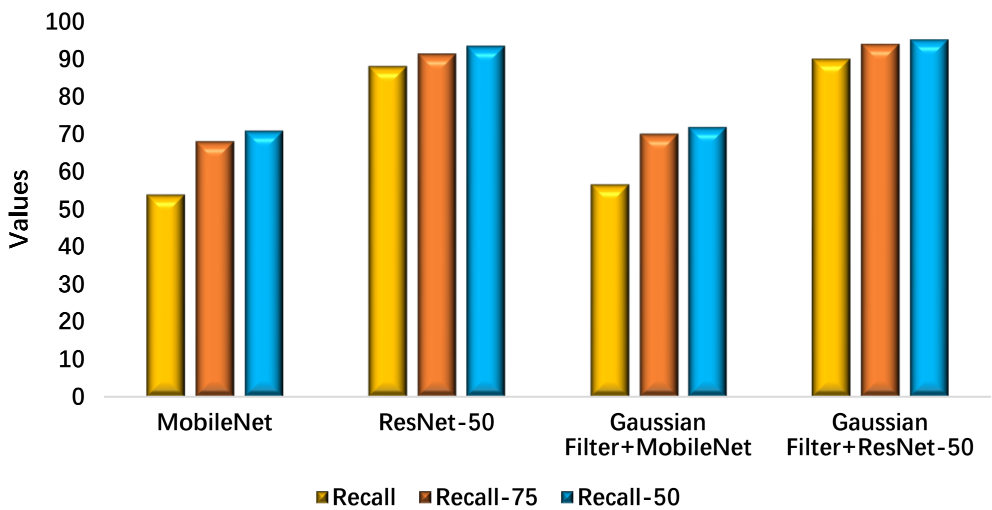

Figure 7. Performance of the employed models based on average recall.

With the modification of the Mask-RCNN by changing the backbone, it was observed that the MobileNet is very easy to train but the performance is not as good as the ResNet-50, in most of the studies carried out using Mask-RCNN, most of the adopted backbones were Resnet101. Looking at the size of our data and the aim of having a model that is lightweight to fit into edge devices and mobile phones, the ResNet-50 can serve as an alternative to ResNet-101 as a backbone, the MobileNet can still be improved to match the performance of the ResNet-50. The Gaussian filter also shows that its effect can improve the performance of the models. Figure 8 shows a test image with the results after detection, Figure 9 shows an example of region proposals and Figure 10 shows segmented pothole.

Figure 8. Predicted pothole with mask.

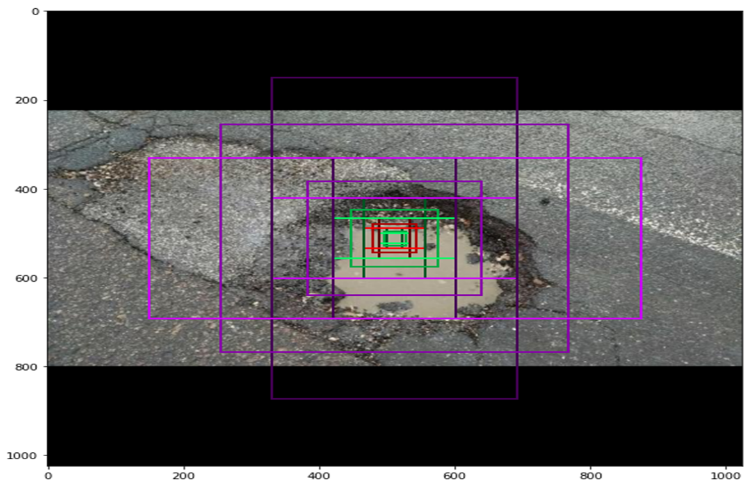

Figure 9. Proposed regions.

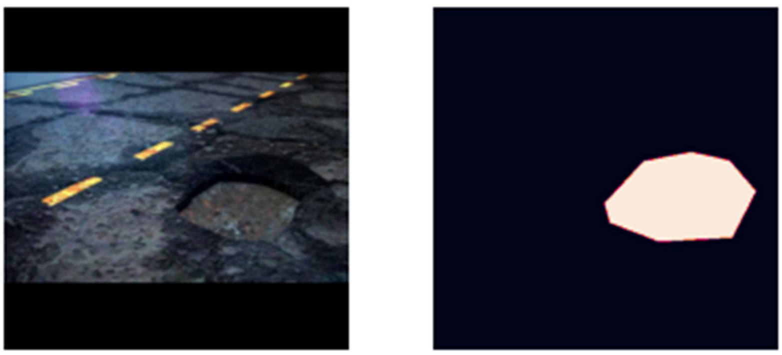

Figure 10. Segmented pothole.

5. Conclusions

As discussed potholes are major causes of accidents in the world, and it is very important to develop a model that will detect potholes efficiently in real-time, Also it is very important to develop models that can fit into edge and mobile devices. In this study Mask-RCNN was employed with different backbones, ResNet-50 and MobileNetv1 to detect potholes, also, to improve the model's performance, Gaussian filtering was applied to the images to reduce noise and improve the visibility of the edges of the potholes, it was observed that the ResNet-50 trained with the smoothening filtered images outperformed all the other models employed in the study, also the techniques proves that the Gaussian filtering can improve the performance of models when it comes to detection of the target object.

So far, the different backbones of ResNet-50 and MobileNetv1employed in this study show an acceptable level of performance in pothole detection and segmentation of potholes, the performance of the models can be increased by using larger datasets. In the future, we will employ different backbones and improved backbone structures to improve performance.

Acknowledgments

We will like to acknowledge the International Research Centre for AI and IoT and the AI and Robotics Institute, Near East University for the provided support in this research.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. V. L. Solanke, D. D. Patil, A. S. Patkar, G. S. Tamrale, A. G. Kale, Analysis of existing road surface on the basis of pothole characteristics, *Global J. Res. Eng.*, **19** (2019).
2. City of San Antonio 311 City Services and Info, Potehole/pavement reapair, 2018. Available from: https://311.sanantonio.gov/kb/docs/articles/transportation/potholes
3. V. Pandey, K. Anand, A. Kalra, A. Gupta, P. P. Roy, B. G. Kim, Enhancing object detection in aerial images, *Math. Biosci. Eng.*, **19** (2022), 7920–7932, https://doi.org/10.3934/mbe.2022370
4. S. M. Hejazi, C. Abhayaratne, Handcrafted localized phase features for human action recognition, *Image Vis. Comput.*, **123** (2022), 104465. https://doi.org/10.1016/j.imavis.2022.104465
5. A. A. Mohamed, F. Alqahtani, A. Shalaby, A. Tolba, Texture classification-based feature processing for violence-based anomaly detection in crowded environments, *Image Vis. Comput.*, **124** (2022), 104465. https://doi.org/10.1016/j.imavis.2022.104488
6. Z. Qu, L. Y. Gao, S. Y. Wang, H. N. Yin, T. M. Yi, An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network, *Image Vis. Comput.*, **125** (2022), 104518. https://doi.org/10.1016/j.imavis.2022.104518
7. K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, Italy, (2020), 2980–2988. https://doi.org/10.1109/ICCV.2017.322

8. A. S. Mubarak, S. Serte, F. Al-Turjman, Z. S. Ameen, M. Ozsoz, Local binary pattern and deep learning feature extraction fusion for COVID-19 detection on computed tomography images, *Expert Syst.*, **39** (2022), e12842. https://doi.org/10.1111/exsy.12842
9. M. Ozsoz, A. Mubarak, Z. Said, R. Aliyu, F. Al Turjman, S. Serte, Deep learning-based feature extraction coupled with multi-class SVM for COVID-19 detection in the IoT era, *Int. J. Nanotechnol.*, **1** (2021). https://doi.org/10.1504/ijnt.2021.10040115
10. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, preprint, arXiv:1804.02767.
11. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Honolulu, USA, (2017), 2261–2269. https://doi.org/10.1109/CVPR.2017.243
12. B. X. Yu, X. Yu, Vibration-based system for pavement condition evaluation, in *Ninth International Conference on Applications of Advanced Technology in Transportation*, (2006), 183–189. https://doi.org/10.1061/40799(213)31
13. K. De Zoysa, C. Keppitiyagama, G. P. Seneviratne, W. W. A. T. Shihan, A public transport system based sensor network for road surface condition monitoring, in *Proceedings of the 2007 workshop on Networked systems for developing regions*, ACM, Kyoto, Japan, (2007), 1–6. https://doi.org/10.1145/1326571.1326585
14. M. B. Sai Ganesh Naik, V. Nirmalrani, Detecting potholes using image processing techniques and real-world footage, in *Cognitive Informatics and Soft Computing*, Springer, (2021), 893–902. https://doi.org/10.1007/978-981-16-1056-1_72
15. L. Huidrom, L. K. Das, S. K. Sud, Method for automated assessment of potholes, cracks and patches from road surface video clips, *Procedia-Soc. Behav. Sci.*, **104** (2013), 312–321. https://doi.org/10.1016/j.sbspro.2013.11.124
16. J. Lin, Y. Liu, Potholes detection based on SVM in the pavement distress image, in *2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, IEEE, Hong Kong, China, (2010), 544–547. https://doi.org/10.1109/DCABES.2010.115
17. M. H. Yousaf, K. Azhar, F. Murtaza, F. Hussain, Visual analysis of asphalt pavement for detection and localization of potholes, *Adv. Eng. Inf.*, **38** (2018), 527–537. https://doi.org/10.1016/j.aei.2018.09.002
18. A. Dhiman, R. Klette, Pothole detection using computer vision and learning, *IEEE Trans. Intell. Transp. Syst.*, **21** (2020), 3536–3550. https://doi.org/10.1109/TITS.2019.2931297
19. S. K. Sharma, S. Mohapatra, R. C. Sharma, S. Alturjman, C. Altrjman, L. Mostarda, et al., Retrofitting existing buildings to improve energy performance, *Sustainability*, **14** (2022), 666. https://doi.org/10.3390/su14020666
20. A. S. Mubarak, Z. S. Ameen, P. Tonga, C. Altrjman, F. Al-Turjman, A framework for pothole detection via the AI-Blockchain integration, in *Lecture Notes on Data Engineering and Communications Technologies*, Springer, (2022), 398–406. https://doi.org/10.1007/978-3-030-99616-1_53
21. J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, H. Balakrishnan, The pothole patrol: Using a mobile sensor network for road surface monitoring, in *Proceedings of the 6th International Conference on Mobile Systems*, ACM, Breckenridge, USA, (2008), 29–39. https://doi.org/10.1145/1378600.1378605

22. A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, L. Selavo, Real time pothole detection using Android smartphones with accelerometers, in *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, IEEE, Barcelona, Spain, (2011), 1–6. https://doi.org/10.1109/DCOSS.2011.5982206
23. X. Yu, E. Salari, Pavement pothole detection and severity measurement using laser imaging, in *2011 IEEE International Conference On Electro/Information Technology*, IEEE, Mankato, USA, (2011), 1–5. https://doi.org/10.1109/EIT.2011.5978573
24. I. Moazzam, K. Kamal, S. Mathavan, S. Usman, M. Rahman, Metrology and visualization of potholes using the microsoft kinect sensor, in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, IEEE, The Hague, Netherlands, (2013), 1284–1291. https://doi.org/10.1109/ITSC.2013.6728408
25. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, USA, (2016), 770–778. https://doi.org/10.1109/CVPR.2016.90
26. C. T. Hendrickson, Applications of advanced technologies in transportation engineering, *J. Transp. Eng.*, **130** (2004), 272–273. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(272)
27. C. Koch, I. Brilakis, Pothole detection in asphalt pavement images, *Adv. Eng. Inf.*, **25** (2011), 507–515. https://doi.org/10.1016/j.aei.2011.01.002
28. M. B. Sai Ganesh Naik, V. Nirmalrani, Detecting potholes using image processing techniques and real-world footage, **1317** (2021), 893–902. https://doi.org/10.1007/978-981-16-1056-1_72
29. Z. Zhang, X. Ai, C. K. Chan, N. Dahnoun, An efficient algorithm for pothole detection using stereo vision, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Florence, Italy, (2014), 564–568. https://doi.org/10.1109/ICASSP.2014.6853659
30. M. Saleh, Z. S. Ameen, C. Altrjman, F. Al-turjman, Computer-vision-based statue detection with gaussian smoothing filter and efficientdet, *Sustainability*, **14** (2022), 11413. https://doi.org/10.3390/su141811413
31. T. Chen, L. Lin, X. Wu, N. Xiao, X. Luo, Learning to segment object candidates via recursive neural networks, *IEEE Trans. Image Process.*, **27** (2018), 5827–5839. https://doi.org/10.1109/TIP.2018.2859025
32. Y. Li, H. Qi, J. Dai, X. Ji, Y. Wei, Fully convolutional instance-aware semantic segmentation, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Honolulu, USA, (2017), 4438–4446,. https://doi.org/10.1109/CVPR.2017.472
33. X. Rong, C. Yi, Y. Tian, Unambiguous scene text segmentation with referring expression comprehension, *IEEE Trans. Image Process.*, **29** (2020), 591–601. https://doi.org/10.1109/TIP.2019.2930176
34. Y. Qiao, M. Truman, S. Sukkarieh, Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming, *Comput. Electron. Agric.*, **165** (2019), 104958. https://doi.org/10.1016/j.compag.2019.104958
35. X. Liu, D. Zhao, W. Jia, W. Ji, C. Ruan, Y. Sun, Cucumber fruits detection in greenhouses based on instance segmentation, *IEEE Access*, **7** (2019), 139635–139642. https://doi.org/10.1109/ACCESS.2019.2942144
36. R. Sagues-Tanco, L. Benages-Pardo, G. Lopez-Nicolas, S. Llorente, Fast synthetic dataset for kitchen object segmentation in deep learning, *IEEE Access*, **8** (2020), 220496–220506. https://doi.org/10.1109/ACCESS.2020.3043256

37. A. M. M. Sizkouhi, M. Aghaei, S. M. Esmailifar, M. R. Mohammadi, F. Grimaccia, Automatic boundary extraction of large-scale photovoltaic plants using a fully convolutional network on aerial imagery, *IEEE J. Photovoltaics*, **10** (2020), 1061–1067. https://doi.org/10.1109/JPHOTOV.2020.2992339
38. Q. Zhang, X. Chang, S. B. Bian, Vehicle-damage-detection segmentation algorithm based on improved mask RCNN, *IEEE Access*, **8** (2020), 6997–7004. https://doi.org/10.1109/ACCESS.2020.2964055
39. T. DeVries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout, preprint, arXiv:1708.04552.
40. F. Song, L. Wu, G. Zheng, X. He, G. Wu, Y. Zhong, Multisize plate detection algorithm based on improved Mask RCNN, in *2020 IEEE International Conference on Smart Internet of Things (SmartIoT)*, IEEE, Beijing, China, (2020), 277–281. https://doi.org/10.1109/SmartIoT49966.2020.00049
41. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Honolulu, USA, (2017), 936–944. https://doi.org/10.1109/CVPR.2017.106
42. L. T. Bienias, J. R. Guillamón, L. H. Nielsen, T. S. Alstrøm, Insights into the behaviour of multi-task deep neural networks for medical image segmentation, in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, Pittsburgh, USA, (2019), 1–6. https://doi.org/10.1109/MLSP.2019.8918753
43. E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683
44. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications, preprint, arXiv:1704.04861.
45. W. Wang, Y. Li, T. Zou, X. Wang, J. You, Y. Luo, A novel image classification approach via dense-MobileNet models, *Mobile Inf. Syst.*, **2020** (2020), 7602384. https://doi.org/10.1155/2020/7602384
46. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Vegas, USA, (2016), 770–778. https://doi.org/10.1109/CVPR.2016.90
47. M. Wang, S. Zheng, X. Li, X. Qin, A new image denoising method based on Gaussian filter, in *2014 International Conference on Information Science, Electronics and Electrical Engineering*, IEEE, Sapporo, Japan, **1** (2014), 163–167. https://doi.org/10.1109/InfoSEEE.2014.6948089
48. A. R. Chitholian, Pothole Dataset, 2020. Available from: https://www.kaggle.com/datasets/chitholian/annotated-potholes-dataset.
49. A. Dutta, A. Zisserman, The VIA annotation software for images, audio and video, in *Proceedings of the 27th ACM International Conference on Multimedia*, ACM, Nice, France, (2019), 2276–2279. https://doi.org/10.1145/3343031.3350535