Pothole Detection

Many datasets used to train artificial intelligence systems to recognize potholes, such as the challenging sequences for autonomous driving (CCSAD) and the Pacific Northwest road (PNW) datasets, do not produce satisfactory results. This is due to the fact that these datasets present complex but realistic scenarios of pothole detection tasks than popularly used datasets that achieve better results but do not effectively represents realistic pothole detection task. In an attempt to improve the detection accuracy of the pothole object detection problems, researchers have proposed varieties of object detection methods enhanced with super-resolution (SR) techniques that are employed to generate an enhanced image from a low-resolution image before performing object detection.

1. Introduction

1.1. Problem Description and Motivation

There are many applications for the detection of objects on the road, with some of the most promising occurring in autonomous driving [1][2], and surface defects that need to be reported to road repair ministries [3][4]. These applications are made possible through cameras that are mounted on moving vehicles. In order to address the challenges of detecting potholes from images and videos, there have been many methods proposed. These methods include processing images or videos captured with cameras from mobile phones [3], unmanned aerial vehicles (UAVs), and drones [5]. However, the methods do not reflect how pothole detection as an object detection problem can be perceived. **Figure 1** shows on the left an image that was taken close up while standing over a pothole. This typically represents how most pothole datasets acquire data and what the state-of-the-art methods have used to train models to detect potholes. The image on the right shows a more realistic scenario of pothole instances captured from a moving vehicle, representing how pothole detection tasks should be perceived. When the methods are presented in a manner that reflects the problem well, the detection performance is not so good because of cases where the amount of noise present in images or videos, most often at low resolution, causes small potholes to appear as insignificant objects that blend into the background. The datasets that present realistic representations of pothole detection problems include PNW [6] and CCSAD [7].

Figure 1. A more realistic task of pothole detection presented in the right column compared to an unrealistic instance.

When evaluating object detection methods' performance, researchers use datasets, such as ImageNet [8] and Microsoft Common Objects in Context (COCO) [9], containing objects that are relatively easy to detect. In addition, the objects often appear large in the images. However, some other objects captured from a distance often appear small, sometimes blending in with the background, and can be challenging to detect using popular object detectors [10]. For images

containing these types of objects to be detected, researchers have found that high-resolution (HR) images offer more input features than low-resolution (LR) images as a result of the lack of input features for small objects [11][12][13].

In an attempt to improve the detection accuracy of the pothole object detection problems, researchers have proposed varieties of object detection methods [14][15][16][17][18][19] enhanced with super-resolution (SR) techniques that are employed to generate an enhanced image from a low-resolution image before performing object detection. In the field of remote sensing, where images are captured from a satellite and most often present the small object detection problem, several methods have been proposed based on super-resolution as well. SR techniques based on convolutional neural networks (CNN), such as single-image super-resolution convolution networks (SRCNN) [14] and accurate image super-resolution using very deep convolutional networks (VDSR) [15], have been proposed and show remarkable results in generating HR images and performing object detection. In addition to CNN-based methods, methods based on generative adversarial network (GAN) [16] have also been proposed. Super-resolution generative adversarial networks (SRGAN) [17], enhanced super-resolution generative adversarial networks (ESRGAN) [18], and end-to-end enhanced super-resolution generative adversarial networks (EESRGAN) [19] have demonstrated better performance in producing both realistic HR images and performing small object detection. These GAN-based models typically consist of generator and discriminator networks that are trained on a pair of LR and HR images, with the generator network generating HR images from the inputted LR images while the discriminator network tries to distinguish the real HR image from the generated HR image. The generator network eventually learns to produce HR images that are indistinguishable from the ground truth HR images, and the discriminator will not be able to distinguish between the images.

Another major challenge in detecting potholes on roads is the cost of the sensor devices used in such a process. Majorly, lidar sensors are exploited for 3D modeling of the surrounding environment to detect obstacles and objects around the vehicle. A single lidar sensor can easily cost thousands of dollars. Cameras have been exploited as cheaper alternatives, but acquiring HR cameras that can capture high-quality images from a moving vehicle can also be expensive.

2. Pothole Object Detection

A variety of devices have been employed to collect data used in road surface anomalies detection. These devices include image acquisition devices, vibration-based sensors, and 3-D depth cameras. Object detection techniques often rely on image data captured by digital cameras [20][21] and depth cameras, thermal imaging technology, and lasers.

To extract the features of a pothole from images, convolutional-neural-network (CNN)-based techniques are more prevalent in this application. These models can accurately model non-linearity in patterns and perform automatic feature extraction on given images. In addition, they are desirable because of their robustness to filtering background noise and low contrast in road images [22]. CNNs have been successfully employed in many applications [1][3][5], but they are not effective in all scenarios. For example, when the object to be detected is small relative to the image, or when high-resolution images are used to mitigate this problem, the computation required to process the data can be prohibitive. This is because CNNs consume a large amount of memory and computation time [23]. To address this, Chen et al. [23] suggest two workarounds to resize input images to the network or using image patches from HR images to train the network. The former workaround is a two-stage system in which a localization network (LCNN) is first employed to locate the pothole instance in the image and then a classification network based on part (PCNN) is utilized to determine the classes.

Salcedo et al. [4] recently proposed a series of deep learning models to develop a road maintenance prioritization system for India. The proposed models include UNet, which employs ResNet34 as the encoder (a neural network subcomponent), EfficientDet, and YOLOv5 on the Indian driving dataset (IDD). Another variation of the you only look once (YOLO) model has also been employed for the task of pothole detection. In a study by Silva et al. [24], the YOLOv4 algorithm was used to detect road damage on a custom dataset that provides an aerial view of roads from a flying drone. The accuracy of the YOLOv4 algorithm and its applicability in the context of identifying damages on highway roads was experimentally evaluated, with an accuracy of 95%.

Asphalt roads can be evaluated by creating 3D crack segmentation models. Guan et al. [25] employed a modified U-net architecture featuring a depth-wise separable convolution in an attempt to reduce the computational workload when working on a multi-view stereo imaging system that contains color images, depth images, and color-depth overlapped images of asphalt roads. The architecture produces a 3D crack segmentation model that considerably outperforms the benchmark models regarding both inference speed and accuracy.

Fan et al. [26] argued that approaches that have employed CNNs for road potholes are faced with challenges of annotating data to be used for training since deep learning models require a large amount of data. The authors thereby proposed a

stereo vision-based road pothole detection dataset and an algorithm that is used to distinguish between damaged road and undamaged roads. The algorithm proposed derived inspiration from graph neural network, where the authors employed an additional CNN layer called the graph attention layer (GAL) to provide optimization for image feature representations for semantic segmentation.

Other methods besides deep learning—such as support vector machines (SVM) and nonlinear SVM—have been explored for extracting potholes from images. Gao et al. [27] employed texture features from grayscale images to train an SVM classifier to distinguish road potholes from cracks in the pavement.

In addition to the aforementioned machine-learning-based techniques, other approaches have been developed. Penghui et al. [28] used morphological processing in conjunction with geometric features from pavement images to detect pothole edges. Koch et al. [29] used histogram shape-based thresholding to detect defective regions in road surface images and subsequently applied morphological thinning and elliptic regression to deduce pothole shapes; texture features within these shapes were compared with those from surrounding non-pothole areas to determine if an actual pothole was present.

As previously mentioned, these proposed techniques produce good results on the test set, but they have not been trained and tested on realistic datasets of high complexity, such as those encountered in autonomous vehicles and unmanned aerial vehicles. Such models will likely underperform when applied to real-world scenarios.

3. Super-Resolution Techniques

Small object detection is commonly exploited in the remote sensing field, where researchers are often faced with small objects in the object categories, making the detection of these objects by state-of-the-art detectors challenging. As images are scaled down by the generic detectors, such as SSD, Faster R-CNN, etc., the performance is reduced. Therefore, most of the proposed methods that use super-resolution images for small object detection are enormous in this field.

Enhanced deep SR network (EDSR) [30] introduces the idea of performing object detection on SR images in the remote sensing field for some of the popularly used architectures [19][31][32]. The ESRGAN [18] architecture improved on the existing super-resolution GAN networks to provide more realistic SR images. The authors employed residual-in-residual dense block (RRDB) with adversarial and perceptual loss to achieve this. The authors achieved a considerable improvement in a subsequent study regarding real-ESRGAN [33] with the use of only synthetic data with high-order degradation modeling, which were close to the real-world degradations.

SwinIR [34] addressed the issue of small object detection with SR data by proposing a transformer that had three parts: a shallow feature extraction step, a deep feature extraction step, and a high-quality image reconstruction step using the residual Swin transformer blocks (RSTB). This transformer produced good results on the DIV2K dataset and the Flickr2K dataset.

Zhang et al. [35] proposed a model called BSRGAN to address degradation issues of SR models that often affect the performance of such models. They proposed that BSRGAN uses random blue shuffle, down sampling, and noise degradation techniques to produce a more realistic degradation of LR images.

The dual regression network (DRN) [36] mapped LR images to HR ones and provided a corresponding degradation mapping function. The authors also found that their method achieved better performance in terms of PSNR (peak signal-to-noise ratio) and the number of parameters.

NLSN for non-local sparse network [37] uses a non-local sparse attention (NLSA) to address the problem of image SR. The method divides the input into hash buckets that contain relevant features, which prevents the network from providing noise or attention to areas of the image with less information during training.

4. Super-Resolution Based Object Detectors

For object detection tasks, both training and inference are affected by the size of the objects. The existing detectors work well with medium-to-large-sized objects but struggle when detecting small-sized objects (objects occupying less than 5% of the overall image size or objects with dimensions in a few pixels). This is because small objects are often indistinguishable from the features of other classes or the background, thereby leading to lower accuracy for the detector.

One technique for improving detector accuracy has been to use data augmentation to oversample small objects of interest, thus increasing the possibility that the small objects will overlap with the prediction [38]. However, this technique has proven to decrease accuracy on other objects in the dataset by reducing the overall amount of training data available for those objects. Another technique proposed for improving detector accuracy is training on both small and large objects of multiple resolutions [39].

YOLOv3 [40] is an object detection system that uses the feature pyramid network (FPN) to quickly provide users with the location of objects in a specific field of view. The system has had great success at detecting small objects due to its ability to detect and locate them without having to perform multiple scans of the same area. One of the significant improvements this network provides is the addition of a new classifier that enables the system to track objects during different stages of their movements, which allows YOLOv3 to locate smaller objects more effectively. However, the network lacks significantly when it comes to processing time. To further improve the performance of small object detection, different modifications have been made to the architecture.

To further improve the performance of YOLOv3 on small object detection and processing speed, Chang et al. [41] proposed amendments to the structure of the network. First, the authors proposed using the K-means algorithm using the width and height of the object's bounding box to obtain appropriate anchor boxes for the objects of interest in a dataset to mitigate the challenge of the objects having different sizes. This modification provides faster network training since the generated anchor boxes are now much closer to the dataset objects.

Lv et al. [42] proposed optimizing the loss function of the YOLOv3 by changing the default loss function L2 and classification loss cross-entropy to GIoU (generalized intersection-over-union) loss function and focal loss, respectively, due to the lack of robustness of the L2 loss function and vulnerabilities, such as the model being sensitive to examples with significant errors and, while trying to adjust, sacrificing example values, with small mistakes. To this end, the GIoU loss function, a variation of the IoU loss function, is proposed to provide a general improvement for the YOLOv3 network

In studies by Bashir and Wang [11] and Courtrai et al. [43], SR networks were used to increase the spatial resolution of LR datasets before feeding the SR images to detector networks for actual detection tasks. Such SR networks have been exploited in recent studies to scale LR images for 2× and 4× scale factors, resulting in remarkable results. In recent years, image generation models that produce single or a pair of images have been widely used for visual representation. Examples include single-image super-resolution (SISR) [44] using a single input; Ferdous et al. [45] used a generative adversarial network (GAN) to produce SR images and SSD to perform object detection on the images; Rabbi et al. [19] combined ESRGAN [18] and EEGAN [46] to develop their own integrated end-to-end small object detection network; Wang et al. [47] proposed a multi-class cyclic GAN with residual feature aggregation (RFA), which is based on both image SR and object detection. The proposed method replaced conventional residual blocks with RFA-based blocks and concatenated the features of the images to improve the performance of the network.

References

1. Dewangan, D.K.; Sahu, S.P. PotNet: Pothole Detection for Autonomous Vehicle System using Convolutional Neural Network. Electron. Lett. 2020, 57, 53–56.

2. Kavith, R.; Nivetha, S. Pothole and Object Detection for an Autonomous Vehicle Using YOLO. In Proceedings of the 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1585–1589.

3. Patra, S.; Middya, A.I.; Roy, S. PotSpot: Participatory Sensing Based Monitoring System for Pothole Detection using Deep Learning. Multimed. Tools Appl. 2021, 80, 25171–25195.

4. Salcedo, E.; Jaber, M.; Requena Carrión, J. A Novel Road Maintenance Prioritisation System Based on Computer Vision and Crowdsourced Reporting. J. Sens. Actuator Netw. 2022, 11, 15.

5. Junqing, Z.; Jingtao, Z.; Tao, M.; Xiaoming, H.; Weiguang, Z.; Yang, Z. Pavement Distress Detection using Convolutional Neural Networks with Images captured via UAV. Autom. Constr. 2022, 133, 103991.

6. PNW Dataset. Available online: www.youtube.com/watch?v=BQo87tGRM74 (accessed on 23 January 2022).

7. Guzmán, R.; Hayet, J.; Klette, R. Towards Ubiquitous Autonomous Driving: The CCSAD Dataset. In Computer Analysis of Images and Patterns: 14th International Conference, CAIP 2011, Seville, Spain, August 29–31 2011, Proceedings, Part II; Springer: Berlin/Heidelberg, Germany, 2011.

8. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. arXiv 2014, arXiv:1409.0575.

9. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common Objects in Context. arXiv 2015, arXiv:1405.0312.

10. Yang, L.; Peng, S.; Nickolas, W.; Yi, S. A survey and Performance Evaluation of Deep Learning Methods for Small Object Detection. Expert Syst. Appl. 2021, 172, 114602.

11. Bashir, S.M.A.; Wang, Y. Small Object Detection in Remote Sensing Images with Residual Feature Aggregation-Based Super-Resolution and Object Detector Network. Remote Sens. 2021, 13, 1854.

12. Haris, M.; Shakhnarovich, G.; Ukita, N. Task-Driven Super Resolution: Object Detection in Low-Resolution Images. In Neural Information Processing. 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12 2021, Proceedings, Part VI; Mantoro, T., Lee, M., Ayu, M.A., Wong, K.W., Hidayanto, A.N., Eds.; Springer: Cham, Switzerland, 2021; p. 1516.

13. Luo, Y.; Cao, X.; Zhang, J.; Cao, X.; Guo, J.; Shen, H.; Wang, T.; Feng, Q. CE-FPN: Enhancing Channel Information for Object detection. arXiv 2022, arXiv:2103.10643.

14. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 38, 295–307.

15. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

16. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

17. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–27 July 2017.

18. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Computer Vision—ECCV 2018 Workshops: Munich, Germany, September 8–14, 2018, Proceedings, Part III; Springer: Berlin/Heidelberg, Germany, 2019; pp. 63–79.

19. Rabbi, J.; Ray, N.; Schubert, M.; Chowdhury, S.; Chao, D. Small-object Detection in Remote Sensing Images with End-to-end Edge-enhanced GAN and Object Detector Network. Remote Sens. 2020, 12, 1432.

20. Kamal, K.; Mathavan, S.; Zafar, T.; Moazzam, I.; Ali, A.; Ahmad, S.U.; Rahman, M. Performance Assessment of Kinect as a Sensor for Pothole Imaging and Metrology. Int. J. Pavement Eng. 2016, 19, 565–576.

21. Li, S.; Yuan, C.; Liu, D.; Cai, H. Integrated Processing of Image and GPR Data for Automated Pothole Detection. J. Comput. Civ. Eng. 2016, 30, 04016015.

22. Sha, A.; Tong, Z.; Gao, J. Recognition and measurement of pavement disasters based on convolutional neural networks. China J. Highw. Transp. 2018, 31, 1–10.

23. Chen, H.; Yao, M.; Gu, Q. Pothole detection using Location-aware Convolutional Neural Networks. Int. J. Mach. Learn. Cybern. 2020, 11, 899–911.

24. Silva, L.A.; Sanchez San Blas, H.; Peral García, D.; Sales Mendes, A.; Villarubia González, G. An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images. Sensors 2020, 20, 6205.

25. Jinchao, G.; Xu, Y.; Ling, D.; Xiaoyun, C.; Vincent, C.S.; Lee, C.J. Automated Pixel-level Pavement Distress Detection based on Stereo Vision and Deep Learning. Autom. Constr. 2021, 129, 103788.

26. Fan, R.; Wang, H.; Wang, Y.; Liu, M.; Pitas, I. Graph Attention Layer Evolves Semantic Segmentation for Road Pothole Detection: A Benchmark and Algorithms. IEEE Trans. Image Process. 2021, 30, 8144–8154.

27. Gao, M.; Wang, X.; Zhu, S.; Guan, P. Detection and Segmentation of Cement Concrete Pavement Pothole Based on Image Processing Technology. Math. Probl. Eng. 2020, 2020, 1360832.

28. Wang, P.; Hu, Y.; Dai, Y.; Tian, M. Asphalt Pavement Pothole Detection and Segmentation Based on Wavelet Energy Field. Math. Probl. Eng. 2017, 2017, 1604130.

29. Koch, C.; Brilakis, I. Pothole Detection in Asphalt Pavement Images. Adv. Eng. Inf. 2011, 25, 507–515.

30. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–27 July 2017; pp. 136–144.

31. Shermeyer, J.; Van Etten, A. The Effects of Super-resolution on Object Detection Performance in Satellite Imagery. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.

32. Wei, Z.; Liu, Y. Deep Intelligent Neural Network for Medical Geographic Small-target Intelligent Satellite Image Super-resolution. J. Imaging Sci. Technol. 2021, 65, art00008.

33. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 1905–1914.

34. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021.

35. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021.

36. Guo, Y.; Chen, J.; Wang, J.; Chen, Q.; Cao, J.; Deng, Z.; Tan, M. Closed-loop matters: Dual Regression Networks for Single Image Super-resolution. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 18–20 June 2020; pp. 5406–5415.

37. Mei, Y.; Fan, Y.; Zhou, Y. Image Super-Resolution With Non-Local Sparse Attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 3517–3526.

38. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for Small Object Detection. In Proceedings of the 9th International Conference on Advances in Computing and Information Technology, Sydney, Australia, 21–22 December 2019; pp. 119–133.

39. Park, D.; Ramanan, D.; Fowlkes, C. Multiresolution Models for Object Detection. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010; pp. 241–254.

40. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. arXiv 2018, arXiv:1804.02767.

41. Chang, L.; Chen, Y.-T.; Wang, J.-H.; Chang, Y.-L. Modified Yolov3 for Ship Detection with Visible and Infrared Images. Electronics 2022, 11, 739.

42. Lv, N.; Xiao, J.; Qiao, Y. Object Detection Algorithm for Surface Defects Based on a Novel YOLOv3 Model. Processes 2022, 10, 701.

43. Courtrai, L.; Pham, M.T.; Lefèvre, S. Small Object Detection in Remote Sensing Images based on Super-resolution with Auxiliary Generative Adversarial Networks. Remote Sens. 2020, 12, 3152.

44. Hui, Z.; Li, J.; Gao, X.; Wang, X. Progressive Perception-oriented Network for Single Image Super-resolution. Inf. Sci. 2021, 546, 769–786.

45. Ferdous, S.N.; Mostofa, M.; Nasrabadi, N. Super Resolution-assisted Deep Aerial Vehicle Detection. In Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, Baltimore, MD, USA, 15–17 May 2019; p. 1100617.

46. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-Enhanced GAN for Remote Sensing Image Superresolution. IEEE Trans. Geosci. Remote Sens. 2019, 57, 5799–5812.

47. Wang, Y.; Bashir, S.M.A.; Khan, M.; Ullah, Q.; Wang, R.; Song, Y.; Guo, Z.; Niu, Y. Remote Sensing Image Super-resolution and Object Detection: Benchmark and State of the Art. Expert Syst. Appl. 2022, 197, 116793.