samples . After all the ground truths were computed, each cluster center was updated using following equation:

The above two steps were repeatedly until the preset number of cycles or the square error converged to the local optimal solution. The distribution of all the samples in surface mine dataset are shown in Fig6. It shows small-scale objects occupy a sizeable proportion in the dataset.

图片

Figure 6. distribution of samples in the **SurMine** (a) distribution of central coordinates of bounding boxes**.** (b) distribution of width and height of bounding boxes.

For different datasets, it is not clear how many clusters to be used. By computing different number of clusters and computing the mean of maximum IOU between the bounding box and individual anchors, we found the mean IOU between anchor boxes and bounding boxes tends to be stable with the number of clusters increasing. However, with the number of anchor boxes increasing, the number of convolution filters in prediction filters increase linearly. We need to get a balance between speed and performance. So, we manually increased the proportion of anchor boxes in the last prediction head to further improve the performance of small-scale objects for Surface Mine dataset.

Under the efforts of the four prediction heads, our model can achieve a considerable improvement for detection of objects with different scales.

# Experiments

We test the performance of our proposed MSFANet on the accuracy of the detector on KITTI and our surface mine dataset. Our implementation environment is based on Pytorch, running on NVIDIA TITAN RTX GPU.

## Datasets

***Surface mine dataset SurMine****.* To construct a robust object detection system for autonomous driving in surface mine, we collect thousands of images to build our dataset. It contains 9 classes with 7165 training and 3508 test images totally collected in different scenarios. Table I shows the statistics for the individual classes of objects.

TABLE I

Details of classes in our dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group** | **Class** | **ID** | **Count** | **Description** |
| ***movable***  ***object*** | car | *1* | *2532* | *including cars, pickups, vans* |
| *truck* | *2* | *16110* | *mining dump trucks* |
| *lorry* | *3* | *2612* | *long distance trucks* |
| *excavator* | *4* | *2300* | *excavators in different size* |
| *auxiliary* | *5* | *3101* | *auxiliary operation vehicles* |
| *person* | *6* | *973* | *pedestrians and workers* |
| ***Infrastru-cture*** | *sign* | *7* | *5038* | *traffic signs* |
| *warning* | *8* | *17320* | *safety warning signs* |
| *rock* | *9* | *3620* | *rocks on the road* |

We find that the distribution of different kinds of samples, in this dataset, have great differences, as shown in Fig7. For example, the images containing rocks are relatively fewer compared to other classes, because the rocks don’t often appear on free space. However, the detection of rocks are particularly important to prevent the accidents. It brought great challenge for our work.

图片

Figure 7. different objects geometry statistics in **SurMine**

More importantly, our dataset contains more challenges shown in Fig8. For instance, false negative would be appeared due to the rock is similar with ground lighting conditions. Besides, objects with tremendous scale difference (e.g., sign and truck) appear in the same image that could cause poor performance of detection task. Also, small scale objects like warning signs and rockfalls may bring great trouble to our detection task.

***KITTI****.* KITTI is a widely used autonomous-driving datasets. The object detection benchmark consists of three classes: car, cyclist, and pedestrian with 7481 training images and 7518 test images, comprising a total of 80256 labeled objects. All images are color and saved as png. Because there is no ground truth for testing, the training and validation sets are made by randomly splitting the training set in half [2]. We train our model with the training images and evaluated it on validation images.

## Implementation Detail

In surface mine dataset detection experiments, all the model were pretrained on the coco with 100 epochs. Then, training on the SurMine with 150 epochs, the batch size is 16, the momentum set as 0.937 and weight decay are respectively 0.0005. The polynomial decay learning rate scheduling strategy was used with initial learning rate 0.001; In the experiments, we use Mosaic as data augmentation method.

In KITTI detection experiments, all the model were pretrained on the coco with 100 epochs. Then, training on the KITTI with 75 epochs, the batch size is 16, the momentum set as 0.937 and weight decay are respectively 0.0005. The polynomial decay learning rate scheduling strategy was used with initial learning rate 0.001.

# CONCLUSION

In this paper, we propose to solve multi-scale, especially small-scale, and hard detected problems in object detection through the guidance of multi-scale feature fusion method, multi-scale attention method and multi-branch prediction heads method. It is an entirely new method in the literature, which is simultaneously take multi-scale, especially the small-scale, and hard samples into account. Based on it we develop a practical object detector, namely MSFANet. Extensive experiments and analyses on surface mine dataset and KITTI further validate its effectiveness, compatibility and efficiency. We combine this detection model with 3D object detection and successfully apply this work to autonomous driving in surface mine.

References

1. Xiangyu Zhang, Xinyu Zhou, Mengxiao, Lin Jian Sun, “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2018, pp. 6848-6856.
2. Bichen Wu, Forrest Iandola, Peter H Jin, and Kurt Keutzer. “Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2017, pages 129–137.
3. J. Redmon and A. Farhadi, “YOLOv4: An incremental improvement.” 2020, arXiv:2004.10934. [Online]. Available: <https://arxiv.org/abs/20> 04 .10934
4. Mingxing Tan, Ruoming Pang, Quoc V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.” in Proc. International Conference on Machine Learning, 2019, pages 6105–6114.
5. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation.” in Proc. IEEE Conf. Computer Vision and Pattern Recognition. (CVPR), 2014, pp. 580–587.
6. R. Girshick, “Fast R-CNN.”in Proc. International Conference on Computer Vision. (ICCV), Dec. 2015, pp. 1440–1448.
7. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real- time object detection with region proposal networks.” in Proc. Neural Information Processing Systems. (NeurIPS), 2015, pp. 91–99.
8. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2016, pp. 779–788.
9. J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger.” in Proc. IEEE Conf. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 7263–7271.
10. J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement.” 2018, arXiv:1804.02767. [Online]. Available: http://arxiv.org/abs/1804.-02767
11. A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite.” in Proc. IEEE Conf. Computer Vision and Pattern Recognition. (CVPR), 2012, pp. 3354–3361.
12. T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection.” in Proc. IEEE Conf. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 2117–2125.
13. W. Liu et al., “SSD: Single shot multibox detector.” in Proc. European Conference on Computer Vision. (ECCV), 2016, pp. 21–37.
14. Eduardo Arnold; Omar Y. Al-Jarrah; Mehrdad Dianati; Saber Fallah; David Oxtoby; Alex Mouzakitis, “A Survey on 3D Object Detection Methods for Autonomous Driving Applications,” IEEE Transactions on Intelligent Transportation Systems, Vol. 20, no.10, pp.3782 – 3795, Oct. 2019.
15. Yaodong Cui; Ren Chen; Wenbo Chu; Long Chen; Daxin Tian; Ying Li; Dongpu Cao, “Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review,” IEEE Transactions on Intelligent Transportation Systems, Early Access, pp. 1 – 18, March. 2021.
16. Sanghyun Woo, Jongchan Park, Joon-Young Lee3 and In So Kweon, “CBAM: Convolutional Block Attention Module, ” in Proc. European Conference on Computer Vision. (ECCV), 2018, pp. 3–19.
17. Jie Hu, Li Shen and Gang Sun, “Squeeze-and-Excitation Networks,” in Proc. Computer Vision and Pattern Recognition. (CVPR), June. 2018, pp. 7132–7141.
18. Xiang Li, Wenhai Wang, Xiaolin Hu and Jian Yang, “Selective Kernel Networks,” in Proc. Computer Vision and Pattern Recognition. (CVPR), June. 2019, pp. 510–519.
19. Haifeng Li, Kaijian Qiu, Li Chen, Xiaoming Mei, Liang Hong, and Chao Tao, “SCAttNet: Semantic Segmentation Network With Spatial and Channel Attention Mechanism for High-Resolution Remote Sensing Images,” IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 1-5, April. 2020.
20. Zequn Qin, Pengyi Zhang, Fei Wu, Xi Li , “FcaNet: Frequency Channel Attention Networks.” 2020, arXiv:2012.11879. [Online]. Available: https://arxiv.org/abs/2012.11879
21. Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, Tat-Seng Chua, “SCA-CNN: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning.” n Proc. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 5659–5667.
22. Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh，Yueh-Hua Wu，Ping-Yang Chen, Jun-Wei Hsieh, “CSPNET: A New Backbone That Can Enhance Learning Capability of CNN.” in Proc. Computer Vision and Pattern Recognition. (CVPR), June. 2020, pp. 1571–1580.
23. Mingxing Tan Ruoming Pang Quoc V. Le, “EfficientDet: Scalable and Efficient Object Detection.” in Proc. Computer Vision and Pattern Recognition. (CVPR), June. 2020, pp. 10781–10790.
24. Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng, “Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection.” in Proc. European Conference on Computer Vision. (ECCV), 2018, pp. 3–19.
25. A. Shrivastava, R. Sukthankar, J. Malik, and A. Gupta, “Beyond skip connections: Top-down modulation for object detection.” 2016, arXiv:1612.06851. [Online]. Available: https://arxiv.org/pdf/1612.06851. pdf
26. S. Bell, C. L. Zitnick, K. Bala, and R. B. Girshick. “Inside- outside net: Detecting objects in context with skip pooling and recurrent neural networks.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2016, pp. 2874–2883.
27. Jongchan Park, Sanghyun Woo, Joon-Young Lee, “BAM: Bottleneck Attention Module.” 2018, arXiv:1807.06514v2. [Online]. Available: https://arxiv.org/pdf/1807.06514.pdf.
28. C. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg. “DSSD: Deconvolutional single shot detector.” 2017, arXiv:1701.06659. [Online]. Available: https://arxiv.org/abs/1701.06659
29. Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He, “Aggregated Residual Transformations for Deep Neural Networks.”, in Proc. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 1492–1500.
30. Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, Qinghua Hu, “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2020, pp. 11534–11542.
31. Gao Huang, Zhuang Liu, Laurens van der Maaten, “Densely Connected Convolutional Networks.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 4700-4708.
32. Charles R. Qi, Li Yi, Hao Su, Leonidas J. Guibas, “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.” in Proc. International Conference on Neural Information Processing. (NIPS), 2017, pp. 5105–5114.
33. Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas , “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 652–660.
34. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.” 2017, arXiv:1704.04861. [Online]. Available: https://arxiv.org/abs/1704.04861
35. Bing Han;Yunhao Wang;Zheng Yang;Xinbo Gao, “Small-Scale Pedestrian Detection Based on Deep Neural Network,” IEEE Transactions on Intelligent Transportation Systems, Vol.21, no.7, pp. 3046–3055, July, 2020.
36. Long Chen; Shaobo Lin; Xiankai Lu; Dongpu Cao; Hangbin Wu; Chi Guo; Chun Liu; Fei-Yue Wang, “Deep Neural Network Based Vehicle and Pedestrian Detection for Autonomous Driving: A Survey.” IEEE Transactions on Intelligent Transportation Systems , Vol.22, no.6, 3234 – 3246, June 2021.
37. Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun, “Deep Residual Learning for Image Recognition.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2016, pp. 770–778.
38. Amir Gholami, Kiseok Kwon, Bichen Wu, Zizheng Tai, Xiangyu Yue, Peter Jin, Sicheng Zhao, Kurt Keutzer, “SqueezeNext: Hardware-Aware Neural Network Design.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2018, pp. 1638-1647.
39. Songtao Liu, Di Huang, Yunhong Wang, “Learning Spatial Fusion for Single-Shot Object Detection.” 2018, arXiv:1911.09516. [Online]. Available: https://arxiv.org/abs/1911.09516
40. Qijie Zhao, Tao Sheng, Yongtao Wang, Zhi Tang, Ying Chen, Ling Cai, Haibin Ling, “M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network.” in Proc. Association for the Advancement of Artificial Intelligence. (AAAI), 2019, pp. 9259–9266
41. Gao Huang, Zhuang Liu, Laurens van der Maaten, “Densely Connected Convolutional Networks.” in Proc. Computer Vision and Pattern Recognition. (CVPR), 2017, pp. 4700-4708.

**First A. Author** **Ruiqi Song** received the M.S. degree from BeiHang University, Beijing, China, in 2016. He is currently a Research associate at The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences， China.

His research interests include automated driving, vision sensing and perception, and machine learning.



**Second B. Yunfeng Ai** received the B.S. degree from Shandong University, Jinan, China, in 2001 and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2006. He was a visiting scholar from December 2015 to October 2016 and postdoctoral researcher from November 2016 to April 2017 at Carnegie Mellon University. He is currently an associate professor of University of Chinese Academy of Sciences.

His current research interests include computer vision, machine learning, robots, and automated driving.

**Third C. Bin Tian** received the B.S. degree from Shandong University, Jinan, China, in 2009 and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2014. He is currently an Associate Professor of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences.

His research interests include automated driving, vision sensing and perception, and machine learning.

**Fourth D. Long Chen** received the Ph.D. degree in signal and information processing from Wuhan University, Wuhan China, in 2013. From October 2010 to November 2012, he was co-trained Ph.D. Student at National University of Singapore. He is currently an Associate Professor with School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. He received the IEEE Vehicular Technology Society 2018 Best Land Transportation Paper Award, IEEE Intelligent Vehicle Symposium 2018 Best Student Paper Award and Best Workshop Paper Award. His interests include autonomous driving, robotics, artificial intelligence where he has contributed more than 110 publications. He serves as an Associate Editor for IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Intelligent Vehicles.

**Fifth E.** **Fenghua Zhu** (M’10-SM’21) received the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2008. He is currently an Associate Professor with the State Key Laboratory for Management and Control of Complex Systems, China.

His research interests are artificial transportation systems and parallel transportation management systems.

**Sixth F. Fei-Yue Wang** (S’87-M’89-SM’94-F’03) received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, New York in 1990. He joined the University of Arizona in 1990 and became a professor and Director of the Robotics and Automation Laboratory (RAL) and Program in Advanced Research for Complex Systems (PARCS). In 1999, he founded the Intelligent Control and Systems Engineering Center at the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Oversea Chinese Talents Program from the State Planning Council and “100 Talent Program” from CAS, and in 2002, was appointed as the Director of the Key Laboratory of Complex Systems and Intelligence Science, CAS. In 2011, he became the State Specially Appointed Expert and the Director of the State Key Laboratory of Management and Control for Complex Systems. Dr. Wang’s current research focuses on methods and applications for parallel systems, social computing, and knowledge automation. He was the Founding Editor-in-Chief of the International Journal of Intelligent Control and Systems (1995−2000), Founding EiC of IEEE ITS Magazine (2006−2007), EiC of IEEE Intelligent Systems (2009−2012), and EiC of IEEE Transactions on ITS (2009−2016). Currently he is EiC of China’s Journal of Command and Control. Since 1997, he has served as General or Program Chair of more than 20 IEEE, INFORMS, ACM, ASME conferences. He was the President of IEEE ITS Society (2005−2007), Chinese Association for Science and Technology (CAST, USA) in 2005, the American Zhu Kezhen Education Foundation (2007−2008), and the Vice President of the ACM China Council (2010−2011). Since 2008, he is the Vice President and Secretary General of Chinese Association of Automation. Dr. Wang is elected Fellow of IEEE, INCOSE, IFAC, ASME, and AAAS. In 2007, he received the 2nd Class National Prize in Natural Sciences of China and awarded the Outstanding Scientist by ACM for his work in intelligent control and social computing. He received IEEE ITS Outstanding Application and Research Awards in 2009 and 2011, and IEEE SMC Norbert Wiener Award in 2014.

图片

appendix. Comparison of the results of (a) Yolov3, (b) Yolov4, (c) CSPNet, (d) MSFANet and (e) Ground Truth