LETTER

Loosening Bolts Detection of Bogie Box in Metro Vehicles Based on Deep Learning

Weiwei QI[†a)], **Shubin ZHENG**[†], **Liming LI**[†], ***Nonmembers***, ***and*** **Zhenglong YANG**[†], ***Member***

SUMMARY Bolts in the bogie box of metro vehicles are fasteners which are significant for bogie box structure. Effective loosening bolts detection in early stage can avoid the bolt loss and accident occurrence. Recently, detection methods based on machine vision are developed for bolt loosening. But traditional image processing and machine learning methods have high missed rate and false rate for bolts detection due to the small size and complex background. To address this problem, a loosening bolts defection method based on deep learning is proposed. The proposed method cascades two stages in a coarse-to-fine manner, including location stage based on the Single Shot Multibox Detector (SSD) and the improved SSD sequentially localizing the bogie box and bolts and a semantic segmentation stage with the U-shaped Network (U-Net) to detect the looseness of the bolts. The accuracy and effectiveness of the proposed method are verified with images captured from the Shanghai Metro Line 9. The results show that the proposed method has a higher accuracy in detecting the bolts loosening, which can guarantee the stable operation of the metro vehicles.
key words: *metro vehicles, loosening bolts detection, deep learning, Single Shot Multibox Detector (SSD), U-shaped Network (U-Net)*

1. Introduction

Bolts in the bogie box are key elements of connecting the axel structure of metro vehicles. Due to vibration and repetitive loads in long time operation, the bolts are apt to get loose and impact the operation safety. Loosening bolts diagnostic is an important inspection item in daily maintenance. Traditionally, this task is performed by human visual inspection of the marked line painted on the bolt surface in advance. However, with the massive construction of metro lines in China, manual inspection is low efficient and costly. It is necessary to develop automatic detection methods.

Automatic bolts detection methods mainly include ultrasonic methods [1], piezoelectric-based methods [2], magnetic filed based methods [3] and vibration methods [4]. However, these methods are not suitable to be deployed in metro vehicles. Recently, methods based on machine vision are applied to bolts detection of metro vehicles. The bolt in the image is small and has less sematic feature. It's a small object detection problem with complex background. Traditional image processing methods [5], [6] combined with traditional machine learning algorithms have poor robustness. With the development of deep learning techniques, Conventional Neural Networks (CNNs) algorithms have shown significant advantages in the feature extraction and stronger robustness, so they have been widely applied in object classification and segmentation. CNNs algorithms are also applied to the loosening bolts detection [7], [8].

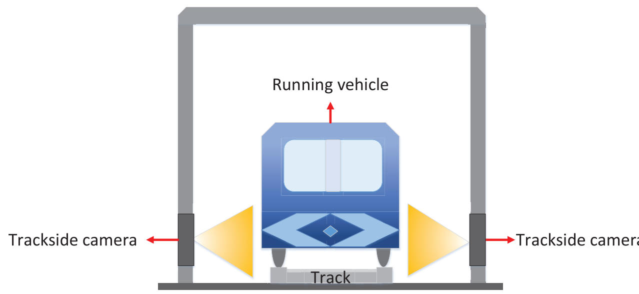

Fig. 1 Sketch map of the trackside image acquisition system installed in the depot

Inspired by the literature [9], we propose a detection method based on CNNs for the loosening bolts detection with images captured by the trackside visual acquisition system installed in depot as shown in Fig. 1. The proposed method consists of the location and segmentation stages. Figure 2 shows the pipeline line of the method. The location stage includes 2 phases. In the first phase, the Single Shot Multibox Detector (SSD) algorithm [10] is adopted to localize the bogie box. Based on the bogie box image cropped from the first localization results, bolts are detected with the improved SSD algorithm in the second phase. In the segmentation stage, the bolts localization results are used to do the semantic segmentation by the U-shaped Network (U-Net) algorithm [11]. The marked line on the bolt is segmented out. Then rotation angle of the marked line is calculated and regarded as the criterion to check whether the bolt is lessening or not.

2. Location Stage Based on SSD

Images are captured by the cameras installed beside the track when the vehicles are moving out of the depot in the early morning, which have low brightness. Besides that, the bogie boxes and some bolts are captured partially into two neighbored images. That will impact the bolt detection accuracy. Image preprocessing are performed before the detection. Gamma transformations are adopted to improve brightness and image combination are performed. The source image size is 2000×4096 and the combined image size is 4000×4096.

Manuscript received May 11, 2022.
Manuscript revised June 25, 2022.
Manuscript publicized July 28, 2022.
[†]The authors are with the Shanghai University of Engineering Science, Shanghai 201620, China.
a) E-mail: 10190002@sues.edu.cn
DOI: 10.1587/transinf.2022EDL8041

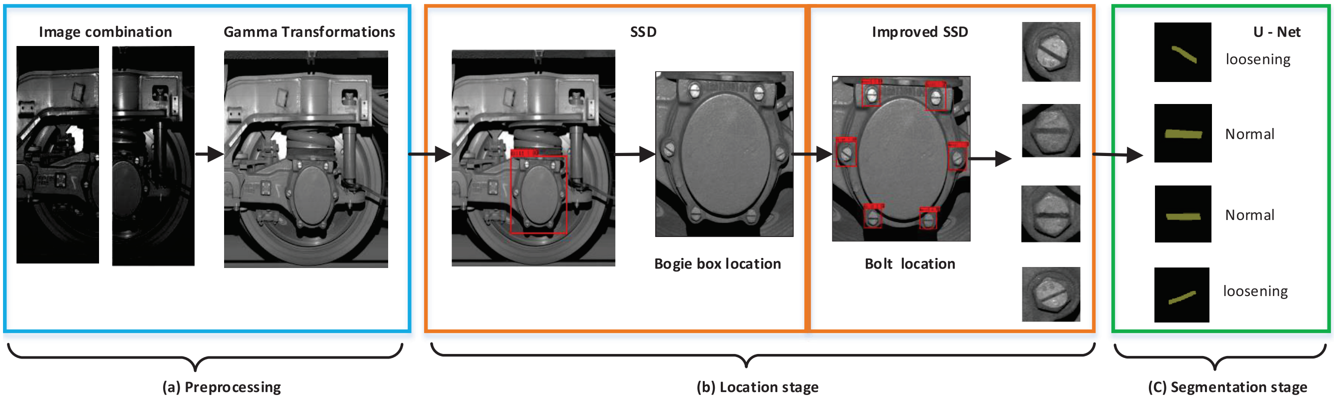

Fig. 2 Pipeline of the proposed method

Both bogie box and bolt location are based on the SSD algorithm. SSD is one of the state-of-the-art object detection algorithms with high accuracy and low time consuming. The detailed structure of the SSD network can be found in [10].

In the bogie box location phase, SSD algorithm with VGG [12] backbone network is adopted directly. Because the bogie box is a middle size object with a typical circular shape in the source image and can be detected easily with high accuracy. SSD has better detection results with lower time consuming than other algorithms. After the bogie box location, the bogie box images are cropped for the bolt location.

In the cropped bogie box image, the bolt is still small object. CNNs algorithms do not perform well in detecting small objects. So We propose an improved SSD algorithm for the bolt location. As illuminated in [13], the convolutional neural network extracts local detailed information in shallow layers and semantic information in deeper layers. Fusion of the local and semantic information can improve the detection performance. As shown in Fig. 3, We improve the SSD network by replacing VGG layers with residual layers constructed with residual units [14], adding feature fusion in shallow layers and reducing the number of default boxes according to the shape of the bolt. The residual block can resolve the problem of gradient vanishing via a short cut connection which also has a function of feature in a unit. We define Res$n$ as a combination of n Res units. Firstly, we add a convolution layer conv 7×7 with normalization and Maxpooling, Res2, Res1 and Res4 layers. Secondly, we add the original extra layers and abandon final layer producing the 1×1 feature map which is contribute less for the bolt detection. Finally, we merge the feature maps from layer Res2 and Res1 to the 38×38 feature map. The 75×75×256 feature map from the Res2 are down-sampled to 38×38×256 with conv 1×1 operation. Then the 38×38×256 feature map is contacted with the 38×38×512 feature map from Res1 to 38×38×768 feature map. We select the feature maps of 38×38×768, 19×19×1024, 5×5×256, 10×10×256, 3×3×256 as the output to the detection module. The bolts have a relatively fixed size. We apply K-means clustering to

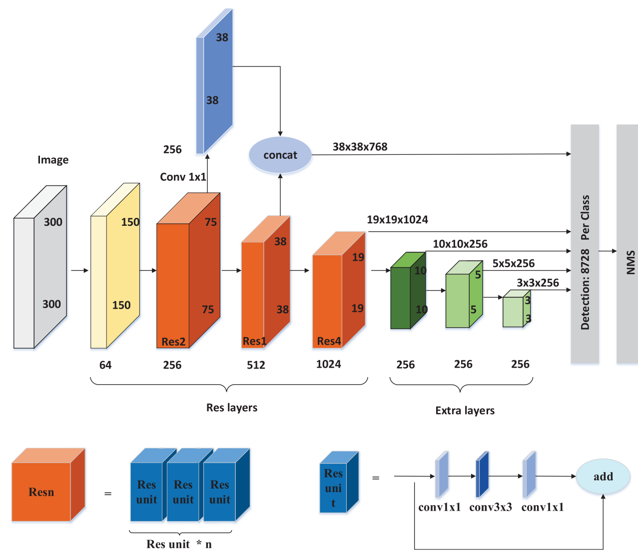

Fig. 3 Improved SSD

determine the size of default bounding boxes. The bolts are classified to 2 clusters of 36×39 pixels and 29×32 pixels. We round the default bounding boxes as (40, 40) and (30, 30). The number of the default bounding boxes of each cell is 2. The total default boxes are reduced from 8732 to 3878. That can reduce the training time.

3. Segmentation Based on U-net

In the segmentation stage, the U-net algorithm is adopted for bolt segmentation. U-net model is a full convolutional network designed for the biomedical image segmentation. U-Net is a good choice for the bolt segmentation because the bolt image and biomedical image have some common characteristics such as simple structure, less color information and less sematic information. We adopt the VGG network as the feature extraction network with 3×3 convolutions and 2×2 max pooling.

After the segmentation, the marked line on the bolt is segmented out. The minimum circumscribed rectangle of the marked line is obtained via the contour extraction algo-

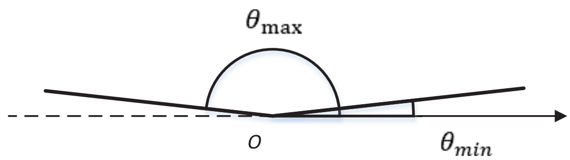

Fig. 4 Angle thresholds

rithm. Then rotation angle $\theta\epsilon[0,\pi]$ of the long side of the rectangle is calculated. We can compare the rotation angle with the predefined angle thresholds $[\theta_{min},\theta_{max}]$ as shown in Fig. 4 to check whether the bolt is loosening or not. The angle thresholds and discriminant function *f(x)* are defined as formulation (1). If *f(x)*=1, the bolt is loosening.

$$f(x)=\begin{cases}0, & (0<\theta<\theta_{min})or(\theta_{max}<\theta<\pi)\\ 1, & (\theta_{min}\le\theta\le\theta_{max})\end{cases} \tag{1}$$

4. Experiments

All algorithms are based on PyTorch framework and run on the Win10 operating system with NVIDIA GeForce RTX 2080. The data set contains pictures from 15 different vehicles of the Shanghai metro line 9 including 720 bogie boxes and 4320 bolts as shown in Table 1.

In the training process of bogie box location, we fine-tune SSD with the model pretrained on VOC dataset with the Adam optimizer. We freeze the VGG layers and train the parameters of extra layers with learning rate =0.0005 for 100 epochs and then unfreeze the VGG layers with learning rate =0.0001 for 100 epochs batch size is set to 8. SSD is compared with the YOlO3, Faster RCNN. According the results in Table 2, all the algorithms have good mean average precision (mAP) but SSD has the best mAP and lowest training time consumption.

In the training process of bolts location, we train the improved SSD from scratch for the hyperparameter including the number of Res units of Res layers, optimizers and learning rate. In the experiments, we find bolts can be detected but there are false positives. False positive can be reduced efficiently by adjusting hyperparameters. We choose the precision and accuracy defined with false positive as the evaluation metrics. Firstly, we compare the detection results with different Res units of Res layers with 256, 512 and 1024 channels respectively as shown in Table 3. With the decreasing of Res units, the false positives are reduced and the precision and accuracy are improved. It verifies that the shallow layers are responsible for detecting small objects. But with too less Res units, the precision and accuracy will decrease due to the poor feature extraction capability. We adopt the configuration of that is Res2, Res1, Res4 with better detection results. Secondly, we compare the converging speed and loss of the RMSprop, Adam, Momentum optimizers and single learning rate. As shown in Fig. 5, the optimizers converge earlier than the single learning rate. Among the 3 optimizers, the RMSprop has the lowest loss. The single learning rate has lower loss than all optimizers. So, we

Table 1 Data set

	Bogie box location	Bolt location	Segmentation
Training sets	420	2400	2400
Validation sets	150	960	960
Testing sets	150	960	960

Table 2 Comparison of methods for the bogie box location

Methods	mAP	Training time(h)	Testing time(s)
Faster RCNN Resnet	96.48%	6.22	120.22
Faster RCNN VGG	95.79%	2.47	119.91
YOLO3	99.31%	2.26	55.45
SSD	99.86%	1.22	61.24

Table 3 Comparison of Res Units for improved SSD

	$Reslayer_{256}, Reslayer_{512}, Reslayer_{1024}$				
	4, 4, 4	4, 2, 4	2, 2, 4	2, 1, 4	1, 2, 4
Precision	96.29%	97.59%	98.33%	99.31%	97.59%
Accuracy	95.58%	96.87%	97.59%	98.55%	96.87%

Table 4 Comparison of methods for bolts location

Methods	mAP	Training time(h)	Testing time(s)
Faster RCNN Resnet	90.77%	34.31	8.58
Faster RCNN VGG	83.52%	14.01	6.82
YOLO3	97.71%	7.83	3.01
SSD	98.21%	5.22	3.32
Improved SSD	99.78%	4.35	3.09

Table 5 Comparison of learning rates for improved SSD

Learning rate	0.00001	0.0001	0.001	0.01
Precision	97.52%	99.85%	86.67%	29.4%
Accuracy	96.79%	99.08%	86.09%	19.61%

Table 6 Comparison of methods for segmentation

Methods	PA	MPA	MIoU	FWIoU	Time(s)
Deeplabv3+	96.18%	93.31%	89.33%	91.79%	0.08
PspNet	96.22%	95.2%	90.57%	92.76%	0.04
U-net	96.89%	95.96%	91.95%	94.01%	0.06

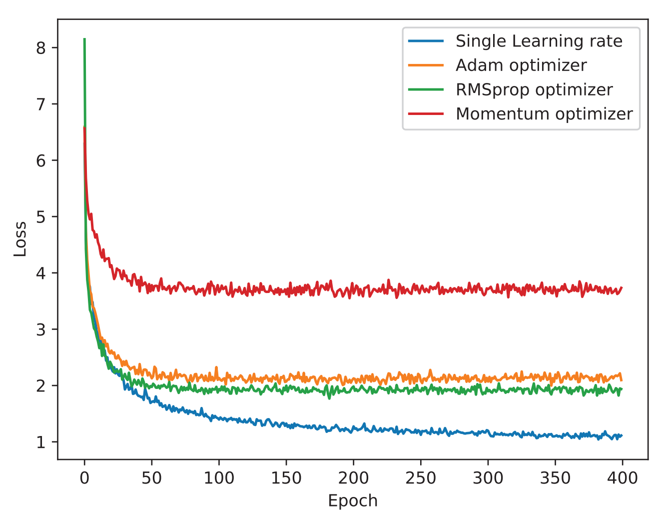

Fig. 5 Comparison of optimizers for improved SSD

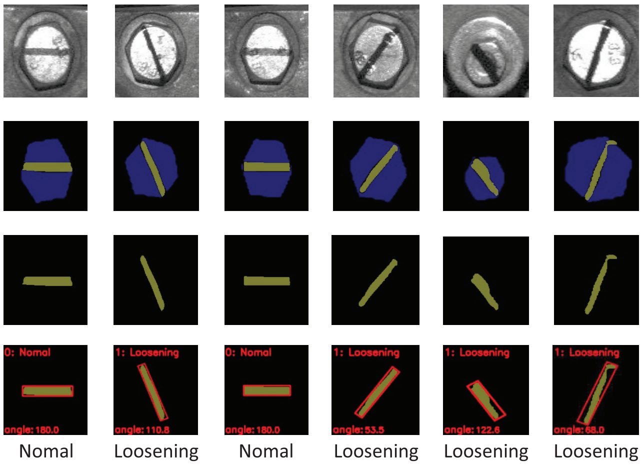

Fig. 6 Loosening bolts detection results

choose the single learning rate to train the network. Finally, we conduct experiment to select the optimal learning rate. Experiment results are shown in Table 4. When the learning rate is set to 0.01, there are many false positives and negatives. The accuracy and precision are improved when the learning rate decrease. When the learning rate is set to 0.0001, there are the minimal false positives. When the learning rate decreases further to 0.00001, the accuracy and precision decrease.

In the segmentation stage, we fine-tune U-Net with the model pertained on VOC dataset with the Adam optimizer. First, we freeze the layers and train the parameters with learning rate for 50 epochs. Then, we unfreeze the layers and train parameters of the network with learning rate for 50 epochs. Based on the configuration of the computer, the batch is set to 2. Unet is verified by comparing with PspNet, deeplabV3+ and Mask RCNN. Pixel accuracy (PA), mean pixel accuracy (MPA), and mean intersection over union (MIoU) are used as evaluation metrics. To compare speed, the time consuming of per image is calculated as shown in Table 6. It is obvious that U-Net has a better performance in accuracy but cost more time than the PspNet. The speed is acceptable because the time has been saved in bolt location stage. After the marked line region is segmented, the minimum circumscribed rectangle of the marked line region is obtained and the rotation angle of the long side of the rectangle is calculated. The results are shown in Fig. 6.

5. Conclusion

This article proposes a detection method of loosening bolts in the bogie box of metro vehicles, which is based on the deep learning algorithms. The proposed method cascades location stage based on SSD and semantic segmentation stage based on U-Net. The experimental results show that the method proposed can effectively detect the loosening bolts. However, there is still room for improvement of bolts location speed and accuracy. In short, the method proposed in this article can effectively detect the loosening bolts and hopes to provide a reference for future loosening bolts detection.

References

[1] T. Stepinski, "Novel instrument for inspecting rock bolt integrity using ultrasonic guided waves," Measurement, vol.177, June 2021.

[2] T. Jiang, Q. Wu, L. Wang, L. Huo, and G. Song, "Monitoring of bolt looseness-induced damage in steel truss arch structure using piezoceramic transducers," IEEE Sensors Journal, vol.18, no.16, pp.6677–6685, Aug. 2018.

[3] K. Hasebe, Y. Wada, and K. Nakamura, "Non-contact bolt axial force measurement based on the deformation of bolt head using quartz crystal resonator and coils," Jpn. J. Appl. Phys., vol.61, no.SG, p.SG1022 (6pp), 2022.

[4] M. Brns, J. Thomsen, S.M. Sah, D. Tcherniak, and A. Fidlin, "Estimating bolt tension from vibrations: Transient features, nonlinearity, and signal processing," Mechanical Systems and Signal Processing, vol.150, Article No. 107224, 2021.

[5] X. Lv, "A novel defect inspection approach using image processing and support vector machines in bolts," 7th Int. Conf. Measuring Technology and Mechatronics Automation IEEE, pp.40–43, 2015.

[6] J.H. Park, T.H. Kim, and J.T. Kim, "Image-based bolt-loosening detection technique of bolt joint in steel bridges," 6th Int. Conf. Advances in Experimental Structural Engineering, 2015.

[7] C.P. Hai, Q.B. Ta, J.T. Kim, D.D. Ho, X.L. Tran, and T.C. Huynh, "Bolt-loosening monitoring framework using an image-based deep learning and graphical model," Sensors (Basel, Switzerland), vol.20, no.12, 2020.

[8] Q.B. Ta and J.T. Kim, "Monitoring of corroded and loosened bolts in steel structures via deep learning and hough transforms," Sensors, vol.20, no.23, Article No. 6888, 2020.

[9] J. Chen, Z. Liu, H. Wang, A. Nunez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," IEEE Trans. Instrum. Meas., vol.67, no.2, pp.257–269, Feb. 2018.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, and A.C. Berg, "Ssd: Single shot multibox detector," Springer, Cham, 2016.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Springer International Publishing, 2015.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Comput. Vis. Pattern Recognit., 2014.

[13] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," Springer, Cham, 2014.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) IEEE, 2016.