Research on improving DeepLabv3+ lightweight network for detection of train tarpaulin damage

Bo Cao*, Changtao Wang
School of Automation and Electrical Engineering, Shenyang Jianzhu University, Shenyang 110168, Liaoning, China

ABSTRACT

In the operation of train pick-up and delivery, there are many problems with human inspection, resulting in potential safety hazards. Considering the wide variety of railway trains, the appearance of the train tarpaulin is selected as the research object, and the damage of the tarpaulin is automatically detected by combining object detection and image segmentation techniques. When using object detection to detect damage, the edge damage detection effect is not ideal, so the image segmentation algorithm is combined with the object detection algorithm, while the traditional segmentation algorithm is not ideal for small damage and narrow and long damage segmentation effects, resulting in low detection accuracy. Therefore, a segmentation algorithm DMV3 is designed, and the DMV3 algorithm is feasible through experiments, which solves the problems of inobvious segmentation and inaccurate detection. Compared with other lightweight segmentation algorithms, it also has great advantages.

Keywords: Tarpaulin train, damage detection, image segmentation, DMV3, MobileNetv3

1. INTRODUCTION

With the increasing volume of railway freight trains, the safe operation of trains has become particularly important. In the field of railway freight, tarpaulin trains, as a common means of transportation, often face various risks and challenges during transportation. Among them, tarpaulin damage is a common but very dangerous problem, which may lead to damage to goods, increase safety hazards, and even accidents.

In 2015, the Sift algorithm was proposed as a cargo inspection method based on a laser scanner and image grey release, in 2018, the Faster-RCNN was proposed to identify the doors and windows, tarpaulins, and foreign objects, and in 2019, the safety status of trucks was proposed based on GoogleNet. In 2020, image processing and Hough Transform were proposed to detect faults. In December 2021, MRPN, linear NMS, and multi-scale RoI pooling were proposed for detection.

Traditional train tarpaulin damage detection usually relies on manual visual inspection, and this method has many limitations: manual inspection is inefficient, prone to omissions, and cannot achieve real-time monitoring. To solve these problems, this paper adds the improved Deeplabv3+ to the damage detection algorithm and combines image processing technology implementation for the automatic detection of train tarpaulin damage.

1.1 MobileNetV3 algorithm architecture

Image input first increases the number of channels through 1×1 convolution; then uses depth convolution in high-dimensional space; then goes through the SE attention mechanism to optimize feature map data; finally goes through 1×1 convolution to decrease the number of channels (using linear activation function). When the step size is equal to 1 and the shape of the input and output feature maps is the same, use the residual to connect the input and output; when the step size (downsampling stage) directly outputs the reduced dimension feature map[1]. The network structure is shown in Figure 1.

*caobotongxue@163.com; phone 18541241910

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 133951U · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049136

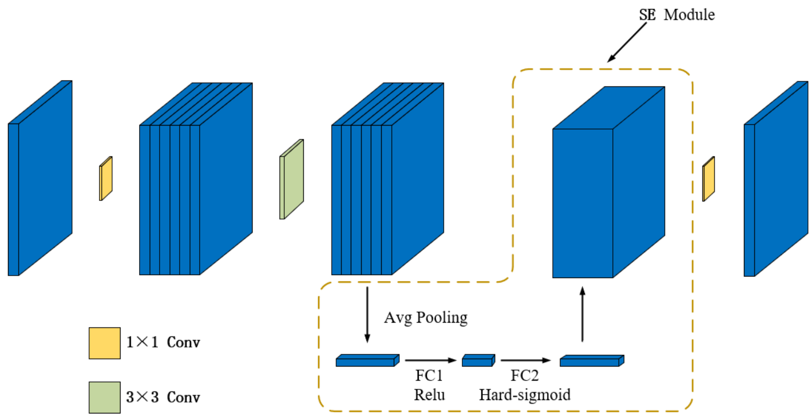


Figure 1. DMV3 Structural diagram.

1.2 Feature

(1) Tail structure: To expand into higher dimensional feature spaces, it is crucial to use 1×1 convolution for the last layer to ensure a rich predictive feature. However, this adds additional computation and latency. To reduce latency while preserving high dimensional features, it was decided to remove the layer before average pooling and use 1×1 convolution to calculate the feature map. With the removal of the feature generation layer, the layer previously used for bottleneck mapping is no longer needed, which reduces processing time and also reduces the number of operands, achieving an increase in speed while reducing the number of operands.

(2) A new activation function is used: replacing swish with h-swish in the structure and replacing sigmoid with h-sigmoid can reduce the amount of computation, improve performance, which is helpful in reasoning speed, and is also very friendly to the quantization process. Using the h-swish activation function can improve the running speed while maintaining accuracy. The formulas are shown in equations (1) and (2).

$$hard-swish = x\frac{ReLU(x+3)}{6} \tag{1}$$

and

$$hard-sigmoid = \frac{ReLU(x+3)}{6} \tag{2}$$

(3) Attention mechanism (SE module): For the obtained feature matrix, each channel is pooled, and then the output vector is obtained through two fully connected layers. In the first fully connected layer, the number of nodes is set to one-quarter of the number of channels in the input feature matrix, while the channels of the second fully connected layer are consistent with the number of channels in the feature matrix. After average pooling and two fully connected layers, the output feature vector can be regarded as a weight relationship analysis for each channel in the feature matrix before the SE. This treatment gives higher weights to channels that are considered more important, and less weight to channels that are considered less important[2].

2. IMPROVED SEGMENTATION ALGORITHM

When identifying the damage to the train tarpaulin, it is impossible to use the traditional target detection method. Therefore, this paper uses an improved segmentation method to show the damage and name it DMV3 and finally identifies the damage through the target detection technology. When using the traditional segmentation method, It will be found that the recognition effect of damage at the edge and elongated damage is not very satisfactory. Therefore, this paper replaces the original backbone network Xception with MobileNetV3[3] based on the original DeepLabV3+, and adds a feature fusion operation to the decoder, which not only improves the recognition effect but also improves the recognition effect. The network is also lighter, greatly improving the speed of operation.

2.1 DMV3 structure

The encoder downsamples the output to a 16× feature map. First, the feature map is sampled up to 4 times by bilinear difference, and then the feature map with 4 times subsampling is spliced. Next, the channels are fused using convolution of 1×1. The feature map with the successful fusion zoom ratio of 4 is upsampled twice, and the feature map with the downsampled multiple of 2 is spliced again. Finally, the channels are fused by 3×3 convolution, and the fused results are up-sampled twice again to obtain the final output feature graph[4].

This improved method can retain more detailed information and has less impact on the main structure of the network, so it can effectively improve the segmentation effect of the tarpaulin. The DMV3 structure is shown in Figure 2.

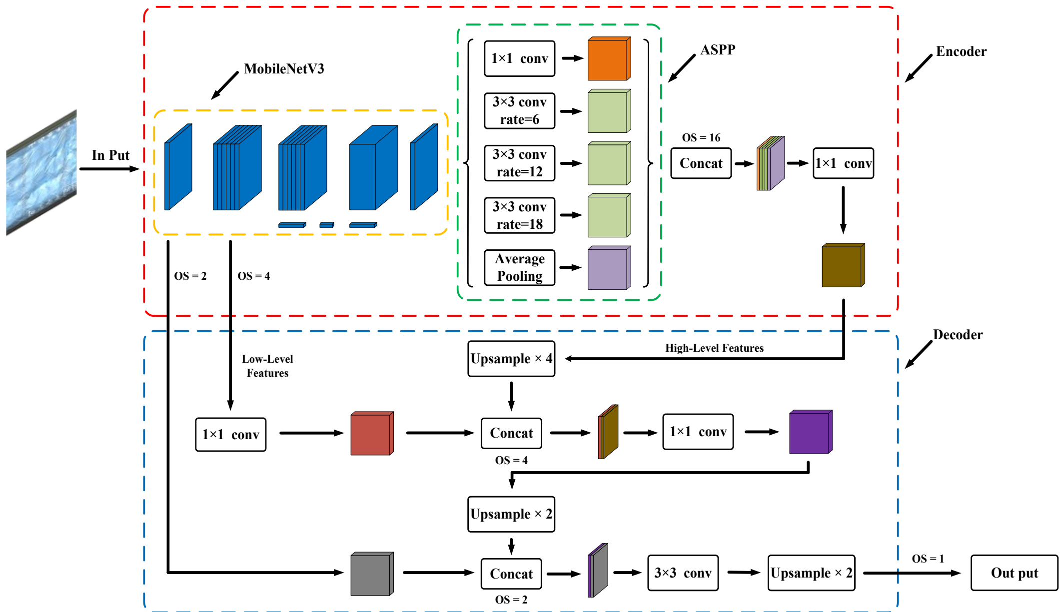


Figure 2. DMV3 Structural diagram.

3. TARPAULIN DAMAGE DETECTION METHOD

Tarpaulin damage is divided into two cases, one is the tarpaulin internal damage, and one is the tarpaulin edge damage, which edge damage detection is more troublesome, so the two cases will be explained separately[5].

3.1 Internal failure detection

After the image is accurately segmented by the DMV3 algorithm, morphological processing is carried out on it. Firstly, the image is preprocessed, the image is grayed and the appropriate structural element (rectangle) is selected. Then, the image is closed (first, the image is expanded to preserve the main area to the greatest extent, and then the image is eroded. Each pixel in the image is manipulated so that it depends on the minimum value of that pixel and its neighborhood. Etching can be used to remove small noise, broken edges, and fine details in the image[6].) Finally, the results of morphological processing, respectively to obtain the tarpaulin area connected domain and damaged area connected domain, by judging whether there is a damaged area in the tarpaulin area, you can know the damage inside the tarpaulin. Figure 3 shows the damage detection process.

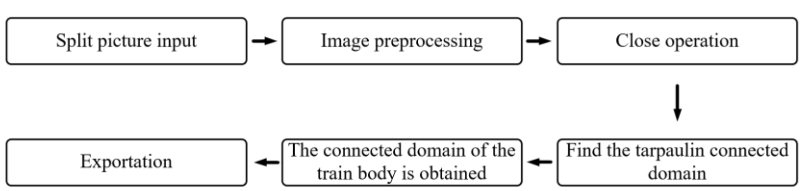


Figure 3. Internal damage detection process.

3.2 Edge breakage detection

Compared with the internal damage, the damage inside the tarpaulin needs to be ignored when obtaining the connected area of the car body and tarpaulin, to avoid interference with the detection of the edge damage, and finally find the outer rectangle of the tarpaulin area, and determine the geometric center coordinates of the outer rectangle. The outer rectangle is a horizontal rectangular box that surrounds all of the tarpaulin area. Figure 4 shows the edge damage detection process.

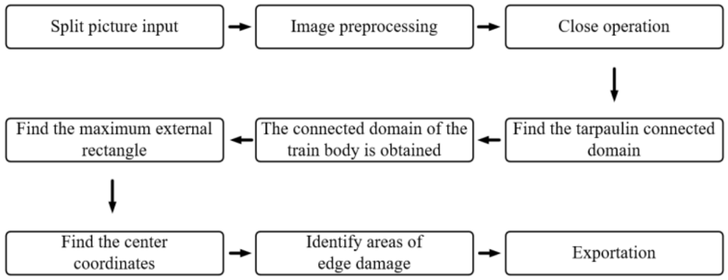


Figure 4. Edge damage detection process.

When edge damage detection is carried out, a scan line will be generated in the center of the image and move down at a fixed distance. In this paper, the distance is set as d=25. The edge of the tarpaulin will intersect with the scan line and produce several intersection points. These intersection points divide the tarp into different areas. A line segment can be obtained by connecting adjacent intersection points, and if the coordinates of the midpoint of the line segment are not in the tarpaulin area, the coordinates of the two endpoints of the line segment are recorded. The scan lines are then moved at the same distance, and if two intersections appear again, it is possible to determine whether they belong to the same damaged area by calculating whether the two line segments intersect. At the same time, the extension length of the damaged area relative to the edge can be estimated by counting the total number of movements of the same damaged area. If the extension length exceeds a certain proportion of the length or width of the rectangular frame, it can be identified as an area of broken edges. Figure 5 is the detection effect diagram.

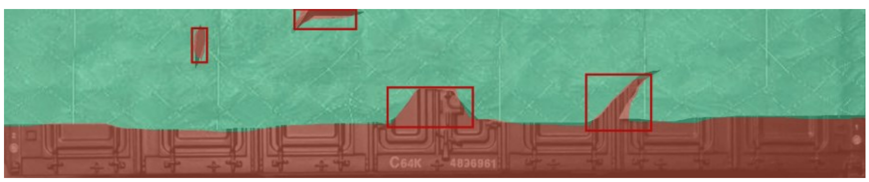


Figure 5. Detection rendering.

4. SIMULATION EXPERIMENT AND ANALYSIS

4.1 Experimental data set

According to the damage points of train tarpaulin, this paper collected 5 kinds of pictures, including severe damage, minor damage, and narrow damage, with a total of 14573 pictures. According to the ratio of 7:3, the data set was divided into training sets and test sets, including 10201 training sets and 4372 test sets.

To verify the segmentation accuracy of the algorithm, the Cityscapes dataset was used, consisting of 2975 high-resolution images in the training set, 500 images in the validation set, and 1525 images in the test set.

4.2 Evaluation index

Average occurring simultaneously (MeanIntersectionOverUnion, MIOU) is the most widely used to measure the application of semantic segmentation accuracy indicators, its value for the training of all kinds of don't at the intersection between predicted values and real values and set the ratio of the average[7]. The formula is shown in equation (3).

$$MIOU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{i=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{3}$$

where k represents the number of sample types except the background (one is tarpaulin, the other is car body). P_{ii} represents the true class I and predicts the correct quantity, P_{ij} represents the number of true class I that was predicted incorrectly, and P_{ji} represents the number of predicted errors for the true class j.

4.3 Verification experiment

Figure 6 is the original image, Figure 7 is the segmentation diagram of deeplabv3+, and Figure 8 is the segmentation diagram of DMV3. According to the comparison between Figures 7 and 8, the segmentation effect of deeplabv3+ is not as good as that of DMV3. It can be seen that the segmentation effect of DMV3 on some relatively small damage and narrow damage at the edge is not very good. When DMV3 is used for segmentation, the segmentation accuracy of small damage and narrow damage at the edge is much higher than that of deeplabv3+, and accurate segmentation can be achieved. Finally, the damage can be detected through the target detection algorithm.

(a) Truck top image

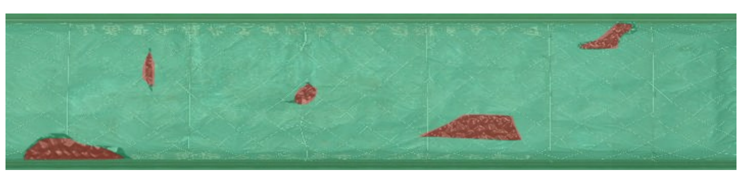
(b) Side image of truck

Figure 6. Original image.

(a) Truck top image

(b) Side image of truck

Figure 7. Deeplabv3+division diagram.



(a) Truck top image

(b) Side image of truck

Figure 8. DMV3 division diagram.

4.4 Comparative experiment and analysis

In the Cityscapes validation set, the proposed algorithm is compared with some popular image semantic segmentation algorithms, and the experimental results are shown in Table 1.

According to the data in Table 1, compared with the SegNet model, the newly proposed algorithm achieved 79.24% more on the Cityscapes validation set, which was 21.29 percentage points higher than SegNet. Compared with DeepLabv3+ models based on Xception and ResNet50, the new algorithm has similar performance, but the number of model parameters is only 2.21 MB, which is far less than other algorithms, showing obvious advantages. Therefore, the new algorithm achieves a good balance between the segmentation accuracy and the number of model parameters.

Table 1. Comparison of performance of different algorithms.

Algorithm	Backbone network	mIoU/%	Parameter quantity/MB
PSPNet[8]	Paper sourced	78.42	84.75
SegNet	Papersource	57.95	29.46
DeepLabv3+	Xception	78.18	78.53
DeepLabv3+	ResNet50	79.20	38.72
DMV3	MobileNetv3	79.24	2.21

Cityscapes validation set is compared with other lightweight image semantic segmentation algorithms, and the experimental results are shown in Table 2.

Table 2. Comparison of lightweight image semantic segmentation algorithms.

Algorithm	mIoU/%	Parameter quantity/MB
FSSNet[9]	62.32	0.20
FastSCNN[10]	69.25	2.33
ENet	60.43	0.37
DMV3	79.24	2.21

As can be seen from Table 2, although the newly proposed algorithm does not have the lowest parameter count, it has the best performance in segmentation accuracy. The number of parameters in the new algorithm is 0.12 MB lower than FastSCNN, but mIoU is 9.99 percentage points higher than FastSCNN, showing a good balance between segmentation accuracy and the number of model parameters.

Due to the replacement of the backbone network with MobileNetv3, reduces the overhead of the image through the backbone network by approximately 7 ms and increases the speed by approximately 11% while reducing the number of operands.

In this experiment, Pytorch3.7.18 deep learning framework was used for training, which was divided into two stages, namely the backbone network freezing stage and the thawing stage. In the freezing stage, the initial learning rate was set to 0.001, the BacthSize was set to 4, and 50 epochs were trained. The initial learning for the thawing phase is set to 0.0001, BatchSize is adjusted to 2, and 100 epochs are trained, for a total of 150 epochs are trained in both phases.

Figure 9 shows the change trend of the loss value during the training process. At the end of the training in the freezing stage, the loss value will increase suddenly, and then the overall trend will decline. It can be seen that at the end of the curve, the loss value fluctuates around 0.12, indicating that the training effect is good.

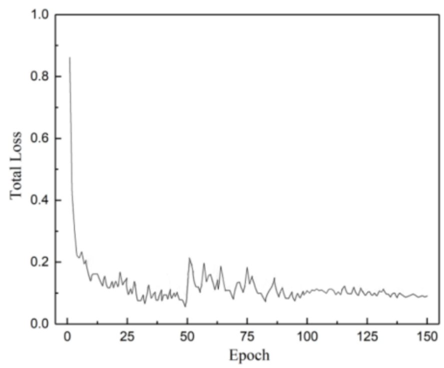


Figure 9. Loss value decline curve.

5. CONCLUSION

(1) Through the experiment, it was found that the DMV3 model could segment the small damage and narrow damage of the train tarpaulin more quickly and accurately.

(2) In the damage detection stage, the morphological image processing method is used, which is combined with the DMV3 segmentation algorithm to greatly improve detection efficiency and accuracy.

(3) Compared with the traditional DeeplabV3+ model, the DMV3 model can detect the damaged area more quickly and accurately, and greatly improve the detection efficiency and accuracy (about 112 ms and 1.5 percentage points). In addition, compared with the mainstream lightweight segmentation network, the DMV3 model can detect the damaged area more quickly and accurately. The improved DeeplabV3+ algorithm can reduce the number of parameters while improving the segmentation accuracy.

REFERENCES

[1] Xiong, Z., Che, W., Bao, Y., et al., "Classification of surface defects in mobilenetv3 hot rolled steel strip based on improvement," Journal of Lanzhou University (Natural Science Edition) 60(02), 182-186 (2024).

[2] Chen, S., Xu, L., Huang, C., et al., "Research on the audio classification of water supply pipeline leakage based on mobilenetv3 convolutional neural network," Journal of Beijing Institute of Technology 7, 1-8 (2024).

[3] Chen, W. and Miao, Z., "Research on urban street view semantic segmentation algorithm based on deeplabv3+algorithm," Wireless Internet Technology 21(09), 103-105 (2024).

[4] Zhao, X., Wang, R., Yang, C., et al., "A natural image semantic segmentation algorithm based on improved deeplabv3," Journal of Shaanxi University of Science and Technology 42(02), 182-188 (2024).

[5] Wang, B. and Asad, S. R., "Solution for sports image classification using modified mobilenetv3 optimized b y modified battle royal optimization algorithm," Heliyon, 9(11), e21603-e21603 (2023).

[6] Said, K. A. M. and Jambek, A. B., "Analysis of image processing using morphological erosion and dilation," Journal of Physics: Conference Series 2071(1), 012033 (2021).

[7] Li, S., Wang, M., Wu, J., et al., "Sea ice detection network for icebreakers in polar environments with attention-based deeplabv3+ architecture," Journal of Physics: Conference Series, 2718 (1), (2024).

[8] Badrinarayanan, V., Kendall, A. and Cipolla, R., "Segnet: a deep convolutional encoder-decoder ar chitecture for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481-2495 (2017).

[9] Zhang, X., Chen, Z., Cai, L., et al., "Fastsemantic segmentation for scene perception," IEEE Trans. Industrial Informatics, 15(2), 1183-1192 (2019).

[10] Poudel, R. P., Liwicki, S. and Cipolla, R., "Fast-SCNN: fast semantic segmentation network," ArXiv, abs/1902.04502, (2019).