Multi-type Rail Region Segmentation Using Context Information Fusion

Xuefeng Ni — Hunan University, Hunan, China, University of Waterloo, ON, Canada
Paul Fieguth — University of Waterloo, ON, Canada
Hongli Liu — Hunan University, Hunan, China
Ziji Ma — Hunan University, Hunan, China
Email: x28ni@uwaterloo.ca, pfieguth@uwaterloo.ca, hongliliu@hnu.edu.cn, zijima@hnu.edu.cn

Abstract

Millions of kilometers of rails need to be inspected throughout the world. Based on vision inspection, rail surface defect detection in a large number of rail images becomes an efficient way to assess rail health status. However there are not only common stock (straight) rails but also many different types associated with rail crossing and connecting, collectively called multi-type rails. Dark rusty regions of rail surfaces and rail bottoms are hard to differentiate in the multi-type rail region segmentation. We discuss how to extract more reliable features to solve this problem by introducing context information fusion. The experimental results show that the context information fusion has a positive effect on multi-type rail region segmentation.

1 Introduction

The technique of rail surface defect detection based on machine vision can be a reliable way to reduce laborious manual inspection and increase the detection efficiency of rail health monitoring [1]. Multi-type rail surface defect detection is an important part of rail health assessment. In detail, multi-type rails include single stock rails, multiple rails, switch rails, wing rails, guard rails, and point rails in turnouts.

In rail surface defect detection, the initial step is to extract rail regions so as to filter unrelated regions, such as fasteners, ballasts, sleepers, and so on. Some researchers have proposed rail extraction methods for single stock rails [2]. Such rails are straight and have similar rectangular shapes. In contrast, the shape, size, number, and orientation of rails are variable in the broader class of images under consideration here, and up to now there is no related research for multi-type rail segmentation, which therefore remains a challenging task in railway inspection.

Semantic segmentation with deep convolutional neural networks (DCNNs) can assign a label for every pixel so as to realize the pixel-level segmentation [3]. In this paper, we regard multi-rail region segmentation as a semantic segmentation task. Features extracted from DCNNs are very helpful for the framework to recognize bright metal areas of rail regions. However, many rail images also have examples of cluttered / poorly illuminated / rusty regions, all of which contribute to significant ambiguity between the desired rail surface, and the unimportant (for inspection purposes) rail bottom. In most cases, the segmentation performance of dark / dirty areas is much poorer than that of bright metal surfaces. Context information is beneficial to explore the relationship between different regions, which can help the framework to learn more abstract global features in the spatial and channel dimensions so as to better differentiate pixels that belong to different classes. Based on the preceding considerations, our proposed research investigates the importance of context information for rail surface region segmentation.

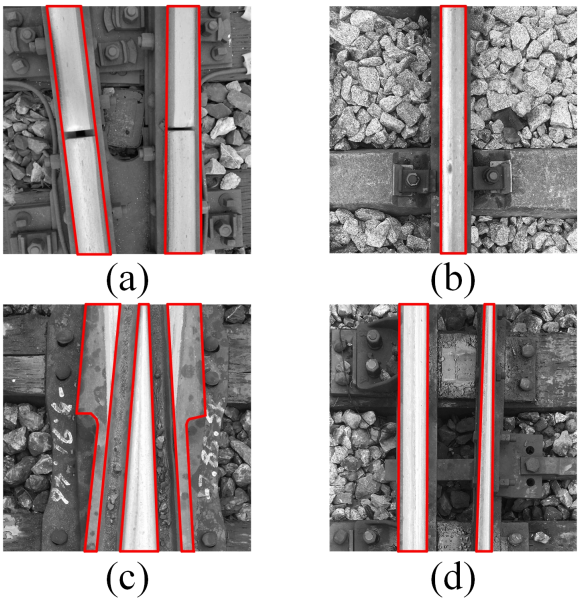

Fig. 1: Difficult parts of multi-type rail region segmentation: (a) Double rails, (b) Single stock rail, (c) Rail frogs, and (d) Switch rails. The red lines indicate the ground-truth contours of rail regions. The variability and ambiguity of the problem can readily be seen.

Although a great deal of work has been undertaken in the domain of rail image segmentation [4] and defect detection [5], nearly *all* of that work has focused on images containing a *single* rail. The single-rail problem is obviously far more constrained and straightforward than multi-type rails, in which images could have one or more rails, rails of different widths, and rails of different shapes and extents.

2 Methodology

An overview of the multi-type rail region segmentation framework is shown in Fig. 2, based on a common encoder-decoder structure. Initially, gray-scale rail images are resized to 512*512 and then their color is normalized. In the encoder part, a backbone with great depth, ResNet101, is utilized for semantic feature extraction. The deep network structure guarantees that high-level features of the global rail regions can be extracted. Subsequently, the feature maps' spatial size is diminished 8 times compared with the input size. After obtaining the high-level feature maps, the position attention module and the channel attention module in DANet [6] are introduced in this framework to model the context-dependence of rail images. The attention mechanism can guide the framework to focus more on important features by weighing operation. More specifically, the position attention module explores the relationship between all pixels so as to obtain spatial context information. The channel attention module emphasizes the interaction between different channels of features. Based on these two kinds of feature interaction, the global context-dependence can be modeled. Therefore, the output attention-weighted feature maps from these modules are fused together by the element-wise addition. Finally, in the decoder part, the convolution filters with upsampling and resizing operations recover the feature maps' size to the original input size so as to predict the pixel-level confidence scores for the segmentation. For the training process, the cross-entropy loss is used for the error regression. Furthermore, in the inference process, in the confidence-score maps, the type of the maximum value in the channel is regarded as the pixel's class.

3 Discussion

We built a dataset with 2757 multi-type rail images, which includes 1791 multiple-rail images, 464 images with single stock rail, 288 images with rail frogs, and 216 images with switch rails. All images are labeled by experts, and there are merely two classes of labels, the rail region, and the background. Then, all images are divided into two parts by the stratified sampling, and each part has the same sample number. Since multiple-rail images' number is much larger than numbers of other types of rail images, in the training set, we augment images with rail frogs and switch rails by 6 times, and then just make data augmentation for images with multiple and

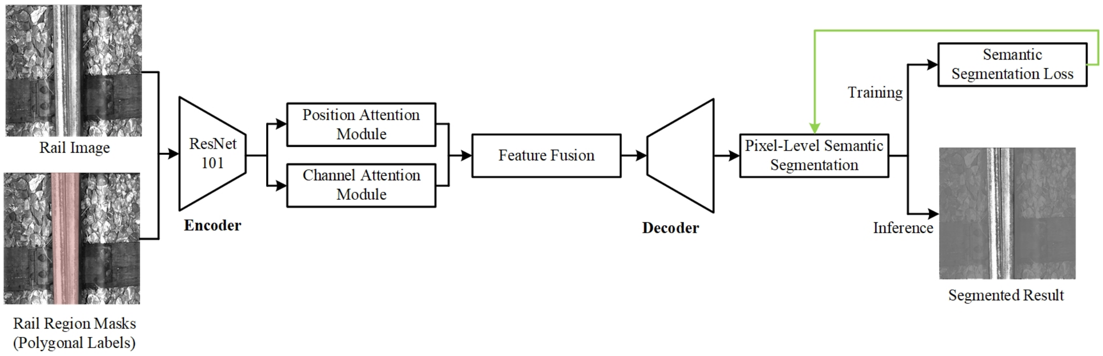

Fig. 2: Overview of the multi-type rail segmentation framework with the position and channel attention module.

single rails by 2 times so as to balance the data number between different classes of images. In summary, the number of the training and test set are separately 3512 and 1378.

The experimental results show that compared with the baseline FCN [7], the framework with context information not only achieves 0.20% improvement in the overall index mean IoU, but also has better performance in segmenting all kinds of rails. Especially, for rail images with large areas of very dark dusty regions, the context information contributes 0.88% IoU improvement for the rail segmentation. The visualization results show that compared with FCN, the segmented regions of dark dusty rails become more complete and some false segmentation on rail regions' boundaries are eliminated after the context information addition. In summary, with the attention modules in DANet, context information and the feature relationship in rail images are explored so as to help the framework differentiate dark dusty regions' types by the relationship differences in the spatial and channel dimension.

In conclusion, the context information is effective in multi-type rail region segmentation. In the future, we plan to explore how to introduce reliable context information to the backbone.

Acknowledgments

The authors gratefully acknowledge the financial support from China Scholarship Council.

References

[1] Q. Li and S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads," *IEEE T. Instrum. Meas.*, vol. 61, no. 8, pp. 2189–2199, Aug 2012.

[2] Q. Li and S. Ren, "A visual detection system for rail surface defects," *IEEE T. Syst. Man. Cy. C.*, vol. 42, no. 6, pp. 1531–1542, Nov 2012.

[3] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE ICCV*, 2015, pp. 1520–1528.

[4] Q. Li, Z. Shi, and et al, "A cyber-enabled visual inspection system for rail corrugation," *Future. Gener. Comp. Sy.*, vol. 79, pp. 374–382, 2018.

[5] D. Zhang, K. Song, and et al, "Two deep learning networks for rail surface defect inspection of limited samples with line-level label," *IEEE T. Ind. Inform.*, 2020.

[6] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on CVPR*, 2019, pp. 3146–3154.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on CVPR*, 2015, pp. 3431–3440.