

Leveraging UAVs for Rapid and Hierarchical Railway Intrusion Detection

Lei Tong¹, Zhipeng Wang¹ (✉), Yong Qin¹, Tian Tang², Bidong Miao³, and Zhaoyu Li³

¹ State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

² Shenzhen United Aircraft Technology Co. Ltd., Shenzhen 518100, China

³ China Railway Electrification Bureau Group Co., Ltd., Beijing 100036, China
zpwang@bjtu.edu.cn

Abstract. The normalized and rapid intrusion detection of the railway operation environment is of great significance for real-time monitoring and early detection of foreign objects that threaten the railway's safe operation. The existing manual patrol and inspection vehicles cannot meet the high demands for rapid detection of some occasional or accidental risks in a short period, such as the intrusion of potential or actual foreign objects. Combining railway segmentation and foreign object detection, this paper proposes a two-stage perception model to achieve hierarchical area-matched intrusion detection and advance the risk assessment of the railway operating environment under the wide field of view of unmanned aerial vehicles (UAVs). With the railway area accurately segmented, the hierarchical division of different risk areas and the foreign objects detection in those hierarchical areas is realized. Leveraging the maneuverability and flexibility of UAVs, the proposed two-stage perception model can effectively improve the efficiency and quality of rapid intrusion inspection of the railway line and surrounding environment.

Keywords: Railway Intrusion Detection, Hierarchical Segmentation, UAV.

1 Introduction

Foreign objects in and around railway areas can pose serious threats to train operations, leading to derailments, damage to rolling stock, and even endangering the lives of passengers. Traditionally, railway intrusion detection relies on mostly manual patrol and inspection vehicles which is time-consuming and labor-intensive with low intelligence levels. Most importantly, they cannot meet the high demands for rapid response to some occasional or accidental emergencies on the lines in operation in a short period. Intrusion detection based on ground video surveillance can only be employed in some limited key railway sections and monitoring points, which means that the full range of the railway line cannot be covered. However, with the advancements in UAV technology, there is an opportunity to revolutionize this process by employing UAV-based solutions, which can possibly provide a hierarchical detection of those foreign objects from a completely different overhead top-down view.

Unmanned aerial vehicles (UAVs) have emerged as a promising technology for a wide range of applications, including power line monitoring [1], vehicle detection [2], structural health monitoring [3], and railway scene surveillance [4], etc. Equipped with various high-precision sensors, the UAV is empowered with a greater ability to collect data from multiple viewing angles to meet diverse inspection needs. The use of UAVs does not need to take up the normal running time of the train while providing a better choice for the automated inspection of the railway operating environment. In light of their potential for enhancing safety and efficiency, UAVs have also gained significant attention in the railway industry to investigate small object detection [4], real-time railway scene parsing [5], railway track detection [6], UAV-LiDAR-based measuring [7], point cloud segmentation [8], ground risk assessment for long-range railway lines [9], etc. A critical aspect of UAV-based railway inspection is the potential or actual foreign object detection of railway operation environment, which, however, has not been researched and discussed adequately.

Moreover, deep learning techniques supported by deep convolutional neural networks have also made prominent advancements in the tasks of object detection [10] and semantic segmentation [11], in which the lightweight and real-time algorithms have more application prospects for the online and onboard perception of the railway operation environment as a result of their excellent potential for deployment on airborne edge computing platforms. The integration of UAV flying platforms and those computer vision algorithms has created a possibility for rapid and hierarchical railway intrusion detection.

This paper investigates UAV-based hierarchical railway intrusion detection with a two-stage algorithm adapted from the lightweight and classical ERFNet [12] and YOLOv5. In the first stage, ERFNet is exploited to fulfill the hierarchical regional division with the railway area as the core area of concern, and YOLOv5 is adopted to detect all potential foreign objects in the whole image. Then in the second stage, hierarchical regions are matched with all those detected potential foreign objects, endowing them with positional information relative to the railway area. In this way, the area-matched potential foreign objects can be divided into differentiated degrees of risk to complete an assessment of the railway operating environment under the wide field of view of UAVs.

2 Methodology

Railway intrusion detection of the railway operation environment is of great significance for the realization of automated UAV-based rapid railway inspections. It is easy to agree that different distances from the railway area represent different degrees of safety risks caused to normal railway operation. This section elaborates on the methodology of distance-based differentiated and hierarchical intrusion detection of potential and actual railway foreign objects in detail.

A general architecture for UAV-based hierarchical railway intrusion detection is shown in Fig. 1. This method first realized the real-time railway segmentation and hierarchical railway area division along the railway line and then completed real-time

object detection. The two steps are both performed from the UAV perspective. The original image taken by the high-resolution camera of UAV is firstly fed into the lightweight segmentation and detection algorithms to obtain an initial potential foreign object detection and railway area segmentation as presented in the detected moving target in (b) and the core area in (c), respectively. Then the core area indicating the railway area is then further processed to obtain hierarchical region divisions, i.e., core area, adjacent area, and peripheral area. After that, all the detected potential or actual foreign objects are matched positionally with the hierarchical region division results to accomplish the hierarchical intrusion detection, as depicted in (d).

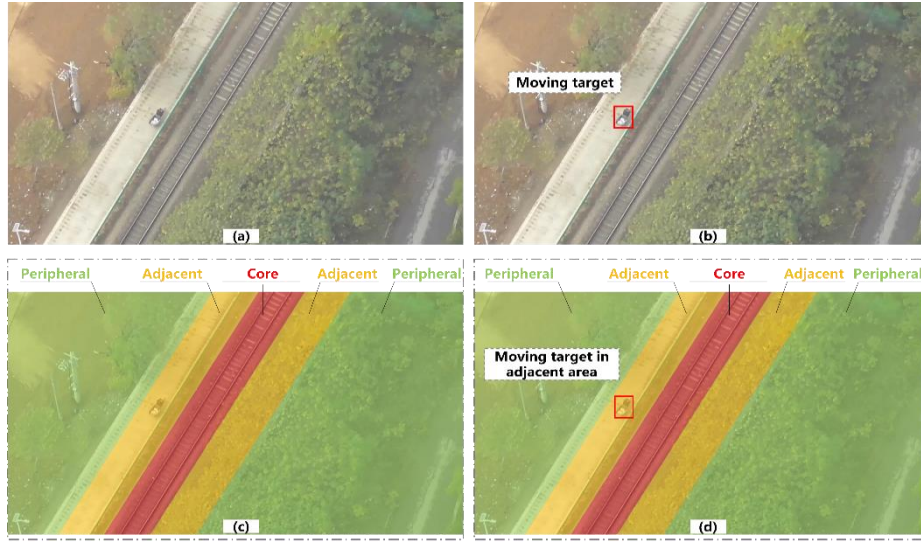


Fig. 1. A general architecture for UAV-based hierarchical railway intrusion detection. (a) The original image; (b) The detected image; (c) The hierarchical railway area division result; (d) The area-matched intrusion detection result.

2.1 Real-time railway segmentation

The remote sensing images collected by UAVs have a wide field of view and rich content, to which the railway area should be paid extra attention. Therefore, in order to obtain regions of different degrees of interest in these images, it is necessary to effectively segment the original image effectively. To achieve the real-time hierarchical perception of the railway operation and surrounding environment, ERFNet is employed to complete the segmentation of the railway area. It is a popular semantic segmentation deep architecture that can make a good tradeoff between high precision and computing efficiency. The superior performance has been verified on platforms of a large-scale GPU card (83+ FPS in a single Titan X) and embedded edge computing device (7 FPS in an NVIDIA Jetson TX1) and public data set Cityscapes.

The network architecture of ERFNet is shown in Fig. 2, in which (a) gives the structure of the main building block Non-bt-1D and (b) gives the encoder-decoder

structure of the network. The block in (a) has the same input and output channel numbers and is a non-bottleneck design, in which all convolutions use one-dimensional kernels with the size of 1×3 or 3×1 . The encoder and the decoder are composed of layers 0-12 and 13-18, respectively. The downsampler is used to downsample feature maps by concatenating the parallel outputs of a single 3×3 convolution with stride 2 and a Max-Pooling module. The upsampling module is realized by simple deconvolution layers with stride 2, which are also called transposed convolutions. The network makes an output with the same resolution as the input image and C channels corresponding to C classes in the dataset.

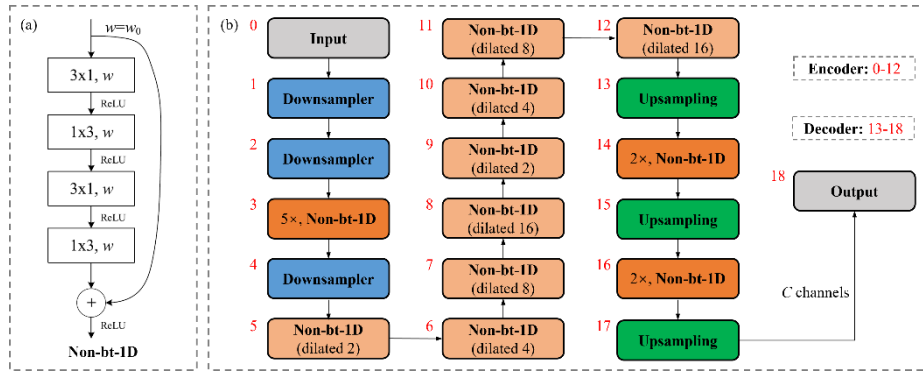


Fig. 2. Structure of residual layer Non-bt-1D and network architecture of ERFNet.

2.2 Hierarchical railway area division

The railway area parsed by the deep networks usually has relatively fuzzy and rough boundaries, which is not conducive to the realization of distance-based hierarchical railway area division. Therefore, the final railway core area needs to be smoothed. Fig. 3 gives three examples that have different shapes of railway area, i.e., (a) trapezoidal railway area, (b) curved railway area, and (c) regular straight railway area. The largest connected area of the initial segmented railway area is denoted as \mathcal{A}_{core}^0 .

Then the center line of the railway area can be sampled uniformly and vertically to \mathcal{N} points, denoted as $\{(x_1, y_1), \dots, (x_{\mathcal{N}}, y_{\mathcal{N}})\}$. Then these \mathcal{N} points are used to perform linear regression to produce a straight centerline. The left and right boundaries of the railway area are determined by moving the centerline to the left and right until it passes through the leftmost and rightmost pixel points of the railway area \mathcal{A}_{core}^0 . The two boundary lines together with the upper and lower boundaries of the image form the smoothed railway area determined as the minimum circumscribed parallelogram (MCP) $\mathcal{A}_{core} = \{(i_c, j_c) | 0 \leq i_c \leq w, 0 \leq j_c \leq h\}$, as given in the middle images of (a), (b), and (c) in Fig. 3. The pixel length of the upper boundary of \mathcal{A}_{core} is defined as d . When moving \mathcal{A}_{core}^0 to the left and right by the distance of λd pixels respectively, the adjacent areas \mathcal{A}_{adj} can be obtained by superimposing them with the image itself \mathcal{I} , as shown in formula (1), in which w and h are the width and the height of the image.

$$\mathcal{A}_{adj} = \{(i_C \pm \lambda d, j_C) | 0 \leq i_C \pm \lambda d \leq w, 0 \leq j_C \leq h, (i_C, j_C) \in \mathcal{A}_{core}\} \quad (1)$$

Here the adjacent area is defined to have the same width as the core area, i.e., λ is taken as 1.0, which can be changed as appropriate. The peripheral areas \mathcal{A}_{per} can be calculated by the formula (2). The final hierarchical area division is presented in the right images of Fig. 3.

$$\mathcal{I} = \mathcal{A}_{core} + \mathcal{A}_{adj} + \mathcal{A}_{per} \quad (2)$$



Fig. 3. Examples of hierarchical railway area division with the MCPs as the core area.

2.3 Area-matched intrusion detection

Potential or actual foreign objects at different distances to the railway area pose different degrees of risk to the safety of railway operations. But it is necessary to first detect all potential foreign objects existing in the whole image. YOLOv5 is a very classic object detection algorithm that has been widely applied and deployed due to its excellent performance. This work also takes YOLOv5 as the detector to detect foreign objects. The locations of detected objects are then matched to the hierarchical areas.

$$Att_{obj} = \begin{cases} \mathcal{A}_{core}, & \text{if } Loc_{obj} \in \mathcal{A}_{core} \ (c_x \in i_C \text{ and } c_y \in j_C) \\ \mathcal{A}_{adj}, & \text{if } Loc_{obj} \in \mathcal{A}_{adj} \ (c_x \in i_A \text{ and } c_y \in j_A) \\ \mathcal{A}_{per}, & \text{if } Loc_{obj} \in \mathcal{A}_{per} \ (c_x \in i_P \text{ and } c_y \in j_P) \end{cases} \quad (3)$$

Assuming that the position of a detected object is $Loc_{obj} = (c_x, c_y)$, then its distance-based area attribute Att_{obj} can be matched by formula (3). Area attributes to which all potential foreign objects belong are assigned. Different area attributes, i.e., core, adjacent, and peripheral area, represent different levels of risk, i.e., high, medium, and low risk, that may be caused to the railway safety operation. Objects assigned \mathcal{A}_{core} are

categorized as high risk of actual foreign objects intruding into the railway area.

3 Experiments

3.1 Data acquisition and datasets

The UAV-based data acquisition, processing, and cloud communication process is shown in Fig. 4(a). Two datasets are built for railway area segmentation and intrusion detection, respectively. With a fixed-wing UAV as shown in Fig. 4(b), the images in the two datasets are collected from the same railway scene in Ma’anshan, Anhui Province, China. The segmentation dataset contains 1602 images in total, including 1442 images for training and 160 images for validation. The dataset has two classes representing railway and non-railway areas respectively. The intrusion detection dataset contains 1123 images in all. The sizes for training and validation are 1010 and 113, respectively. The dataset concerns one class of moving objects, i.e., person. Fig. 4(c) presents two embedded edge computing devices that can be carried on the UAVs. They can be used to run the developed lightweight intelligent algorithms in real-time and complete mutual communication with the cloud.

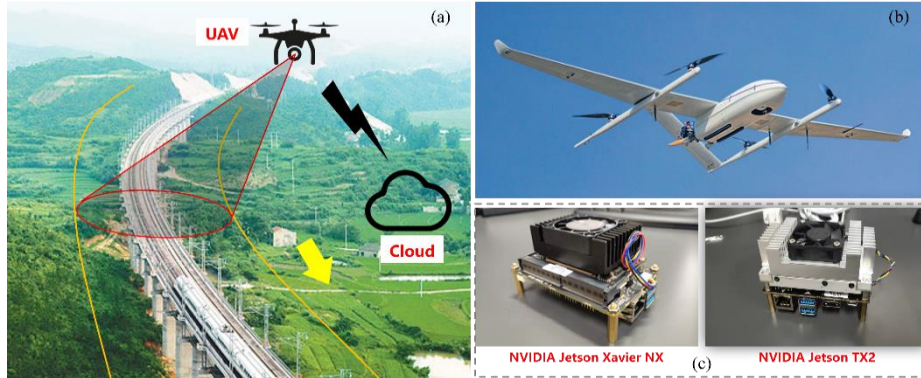


Fig. 4. UAV-based railway operating environment data acquisition and deployment platforms.

3.2 Distance-based area division and object detection

Experiment results on the distance-based area division and potential foreign object detection of the full image are presented. The trained ERFNet obtains a mIoU of 0.956 on the validation dataset. The top row of Fig. 5(a) shows some initial segmentation results by ERFNet, which have rough edges or truncated segments. Then the segmentation results are smoothed by their MCPs to generate a hierarchical division of risk areas, as shown in the bottom row of Fig. 5(a). For the sake of picture clarity, the corresponding peripheral areas are not drawn. As the smallest one in the yolov5 model series, YOLOv5s achieves a mAP@0.5 of 0.921 on the validation dataset. The evaluation P-R curve is given in Fig. 5(b). Figures (c) and (d) of Fig. 5 present some detected visual samples. Despite the high mAP, moving persons on the ground, as

small as a few pixels in a 1080p image, are quite difficult to recognize, even for human eyes. This also illustrates the great challenge of small object detection for UAV aerial images with a large field of view, as shown in the FP and FN cases in Fig. 5(d).

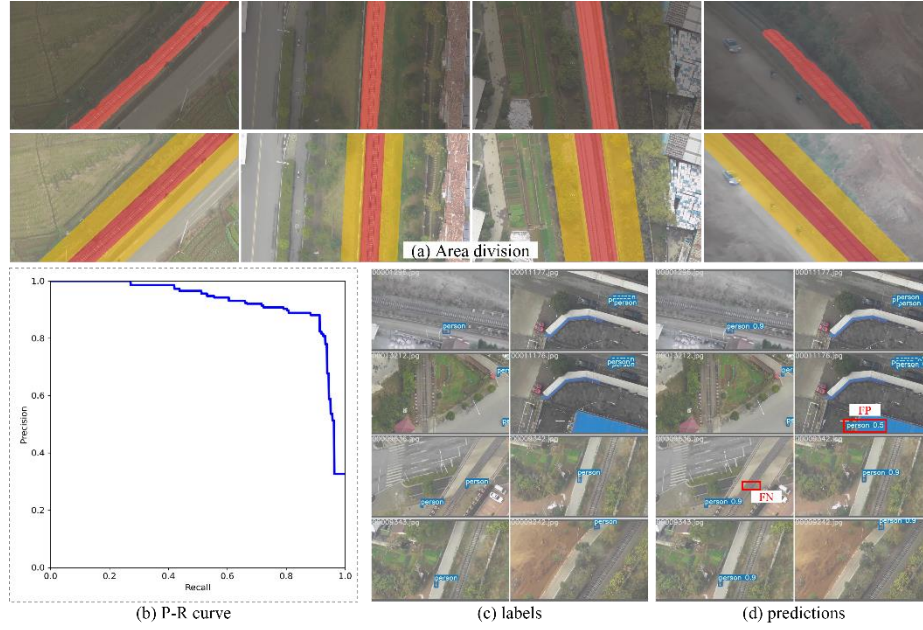


Fig. 5. Distance-based area division (a) by ERFNet and experiment results of YOLOv5s including P-R curve (b) and some object detection visual examples (c) and (d).

3.3 Hierarchical railway intrusion detection

With the results on both area divisions of differentiated risk levels and potential foreign object detection combined, the hierarchical railway intrusion detection is completed. It's verified the proposed MCP smoothing method has excellent effects.



Fig. 6. Some visual samples for UAV-based hierarchical railway intrusion detection results.

As given in Fig. 6, detected potential foreign objects (persons) with different risk levels are represented by circles of different colors, which can be used to assess rail-

way intrusion risks. Red, yellow, and green represent high, medium, and low risks, respectively. For the sake of image clarity, the peripheral areas are not drawn in color.

4 Conclusion

Rapid intrusion detection under the UAV's wide field of view has great advantages for real-time monitoring and early detection compared with other inspection methods. This paper proposed a hierarchical area-matched intrusion detection method based on UAV remote sensing images. Experiments on deep and lightweight segmentation and detection architectures show promising results. Future work will conduct more in-depth research on a large number of small object detection challenges and more refined regional divisions.

Acknowledgment. This work is supported in part by the Fundamental Research Funds for the Central Universities 2022YJS111, in part by the National Key R&D Program of China under Grant 2022YFB4300601, and in part by the National Natural Science Foundation of China under Grant 61833002.

References

1. Foudeh, H.: An advanced unmanned aerial vehicle (UAV) approach via learning-based control for overhead power line monitoring: a comprehensive review. *IEEE Access* 9, 130410-130433 (2021).
2. Bouguettaya, A.: Vehicle detection from UAV imagery with deep learning: a review. *IEEE Trans Neural Netw Learn Syst* 33(11), 6047-6067 (2022).
3. Kang, D.: Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. *Comput-Aided Civ and Inf* 33(10), 885-902 (2018).
4. Li, Y.: Multi-block SSD based on small object detection for UAV railway scene surveillance. *Chinese J. Aeronaut.* 33(6), 1747-1755 (2020).
5. Tong, L.: Fully decoupled residual convNet for real-time railway scene parsing of UAV aerial images. *IEEE Trans on Intell Transp Syst* 23(9), 14806-14819 (2022).
6. Tong, L.: Anchor-adaptive railway track detection from unmanned aerial vehicle images. *Comput-Aided Civ and Inf* (early view), 1-19 (2023).
7. Geng, Y.: UAV-LiDAR-based measuring framework for height and stagger of high-speed railway contact wire. *IEEE Trans on Intell Transp Syst* 23(7), 7587-7600 (2022).
8. Geng, Y.: 3DGraphSeg: a unified graph representation-based point cloud segmentation framework for full-range highspeed railway environments. *IEEE Trans Industr Inform* (early access) (2023).
9. Bertrand, S., Raballand, N.: Ground risk assessment for long-range inspection missions of railways by UAVs. In: *ICUAS 2017*, pp. 1343-1351. Miami, FL, US (2017).
10. Zhao, Z.: Object detection with deep learning: a review. *IEEE Trans Neural Netw Learn Syst* 30(11), 3212-3232 (2019).
11. Mo, Y.: Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* 493, 626-646 (2022).
12. Romera, E.: ERFNet: efficient residual factorized convNet for real-time semantic segmentation. *IEEE Trans Intell Transp Syst* 19(1), 263-272 (2018).