

Anchor-adaptive Railway Track Detection from Unmanned Aerial Vehicle Images

Lei Tong^{1,2} | Limin Jia^{1,2} | Yixuan Geng^{1,2} | Keyan Liu^{1,2} | Yong Qin^{1,2} | Zhipeng Wang^{1,2}

¹State Key Laboratory of Railway Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

²Key Laboratory of Railway Industry of Proactive Safety and Risk Control, Beijing Jiaotong University, Beijing, China

Correspondence

Zhipeng Wang, State Key Laboratory of Railway Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044 China
Email: zpwang@bjtu.edu.cn

Funding information

National Key R&D Program of China, Grant Number: 2022YFB4300601; State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Contract Number: RCS2022ZT005; National Natural Science Foundation of China (Nos. 61833002 and 62033004); Research project of China State Railway Group Co., Ltd (No. P2022X001).

ABSTRACT

Autonomous railway inspection with unmanned aerial vehicles (UAVs) has huge advantages over traditional inspection methods. As a prerequisite for UAV-based autonomous following of railway lines, it is quite essential to develop intelligent railway track detection algorithms. However, there are no existing algorithms currently that can efficiently adapt to the demand for the various forms and changing inclination angles of railway tracks in the UAV aerial images. To address the challenge, this paper proposes a novel anchor-adaptive railway track detection network (ARTNet), which constructs a dual-branch architecture based on projection length discrimination to realize full-angle railway track detection for the UAV aerial images taken from arbitrary viewing angles. Considering the potential capacity imbalance of the two branches that can be caused by the uneven distribution of railway tracks in the dataset, a balanced transpose co-training strategy is proposed to train the two branches coordinately. Moreover, an extra customized transposed consistency loss is designed to guide the training of the network without increasing any computational complexity. A set of experiments have been conducted to verify the feasibility and superiority of the ARTNet. It is demonstrated that our approach can effectively realize full-angle railway track detection and outperform other popular algorithms greatly in terms of both detection accuracy and reasoning efficiency. ARTNet can achieve a mean F1 of 76.12 and run at a speed of 50 more frames per second.

1 INTRODUCTION

The rapid and effective inspection of the infrastructures in the area for train traveling is of great significance to the safe operation of railway systems. Currently, most of the facilities of the railway industry are inspected via human or inspection vehicles (Ying et al., 2014; Gibert et al., 2017; Zhong et al., 2022). The inspection vehicles have to occupy the train operation diagram and thus shorten the running time for other normal trains. Manual inspections are fairly inefficient and labor-intensive for railways extending thousands of miles. Usually, patrol workers and inspection vehicles can only be limited to a very short maintenance time window at night to conduct railway inspections under difficult lighting conditions, especially for high-density lines like the Beijing-Shanghai high-speed railway. Even if supplementary light

devices can be used at night, the quality of inspection based on human visual observation will be greatly compromised. The data collected by optical sensors deployed on the inspection vehicles will also be quite unsatisfactory because of the unavoidable information losses. As a result of the situation, subsequent defect positioning and detection will be adversely affected.

In recent years, unmanned aerial vehicles (UAVs) and deep learning have been developed rapidly and widely applied to the civil infrastructure field (Rafiei and Adeli, 2018; Rafiei and Adeli, 2018) including the railway operation and maintenance filed (Li et al., 2022; Oudshoorn et al., 2022; Wang et al., 2023) and other fields (Martins et al., 2020; Macias-Garcia et al., 2021; Cheng et al., 2023) because of their superior performance. The mounted camera on UAVs is developed to inspect the inaccessible or unreachable areas of

civil structures and facilities effectively. Researchers have taken excellent advantage of UAVs to investigate cracks or damage detection and assessment for bridges (Yeum and Dyke, 2015; Kang and Cha, 2018; Liu et al., 2020) and other civil structures (Jiang and Zhang, 2020), vision-based bridge component recognition (Narazaki et al., 2020) and cable force estimation (Tian et al., 2021), civil structural health monitoring (Kang and Cha, 2018; Xu et al., 2020), structural displacement or deflection measurement and monitoring (Weng et al., 2021; V. Shajihan et al., 2022; Zhuge et al., 2022), 3D reconstruction (Xu and Liu, 2022), construction site monitoring (Bang et al., 2022), and post-disaster damage assessment of buildings (Cheng et al., 2021). UAVs are also widely adopted in the field of power line monitoring (Kim et al., 2020) and the application of this technology is becoming more and more mature (Foudeh et al., 2021). Moreover, with the emergence of payloads with better performance that can integrate a variety of high-precision sensors, UAVs are more empowered to conduct inspections of various key infrastructures with greater potential. The application of UAVs for railway infrastructure inspection also has huge and unique advantages over traditional inspection methods. UAVs can efficiently collect data on railway facilities from railway lines in service without occupying the train operation diagram in a non-contact remote sensing way. Therefore, as an important auxiliary inspection technique other than the traditional inspection methods, UAV-based railway inspection is a crucial research trend for high-speed railway safety operations.

In light of their superior advantages, some researchers have exploited UAVs to investigate: 1) **image based**: railroad track components inspection (Guo et al., 2021), small objects detection of railway scene (Li et al., 2020), rail surface defects detection (Wu et al., 2018; Wu et al., 2022), rail fastener defect inspection (Chen et al., 2019), catenary support device inspection (Liu et al., 2020), railway scene dehazing to enhance the effect of railway object detection (Wu et al., 2020), and railway scene parsing (Tong et al., 2022); 2) **point cloud based**: rail track detection (Sahebdivani et al., 2020), contact wire measurement with LiDAR (Geng et al., 2022) and point cloud segmentation for railway environments (Geng et al., 2023). However, most of the works are based on the data collected by UAVs that are completely controlled by human labor. In practice, manually manipulating for data collection not only spends much time and labor costs but also leads to unexpected missing shots of the key facilities if the attitude of the UAV or the mounted gimbal cannot be adjusted in time and correctly during flight and ultimately leads to missed inspections of key components or facilities. Thus, the data acquisition section has become the critical bottleneck of UAV-based railway inspection.

To realize intelligent UAV-based inspection of the railway track area, UAV-based automatic data collection of this area is necessary. As a prerequisite of that, it is quite essential to develop intelligent railway track detection algorithms. In the data acquisition process with UAVs, they need to know

where the area is and be able to locate the area during the flight process in the first place. On the one hand, railway tracks are the most iconic facility in the area. On the other hand, the gauge of the track pairs always maintains a constant value (1435mm in China), which can provide a very accurate position reference for locating the track area accurately. Hence, concerning the issues discussed above, this paper manages to establish a real-time railway track detection architecture for UAV aerial images of railway scenarios. However, according to our knowledge, the task of railway track detection towards UAV aerial images is seldom investigated. Although some previous works (Guclu et al., 2021; Saini and Singh, 2021) have developed railway track extraction or detection models based on traditional vision recognition algorithms, they still cannot adapt to the aerial images of changing railway scenarios. Rail track detection from UAV point cloud based 3D model construction mentioned before is impossible to produce real-time closed-loop information feedback to the UAV (Sahebdivani, Arefi and Maboudi, 2020). Despite that, lane detection powered by deep learning has made significant progress in recent years with the help of the lane detection benchmarks: TuSimple (TuSimple, 2017), CuLane (Pan et al., 2018), and LLAMAS (Behrendt and Soussan, 2019). In view of the similarities between lane detection and railway track detection, these works provide a possible solution for this issue.

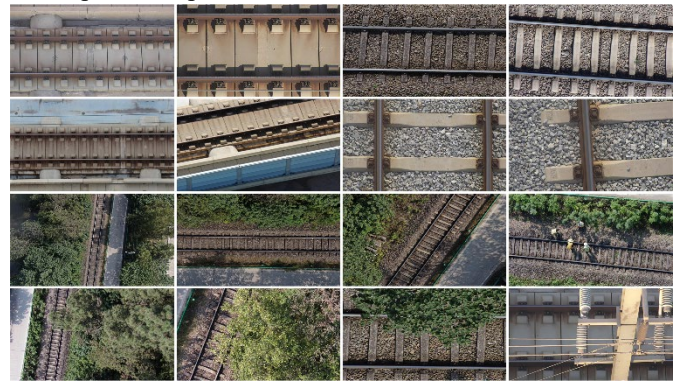


FIGURE 1. Various forms of railway track with changing backgrounds and pixel widths from the perspective of UAVs. Their shapes can be affected by the flying height of the UAV, the attitude of the UAV, and the attitude of the mounted gimbal, etc.

The existing lane detection methods can be divided into three categories according to lane representation strategy: parametric prediction methods, segmentation-based methods, and anchor-based methods. In the parametric prediction methods, the lanes are expressed by curve equations. So, the parameters of the formulated equation are predicted directly corresponding to the lanes in the image. PolyLaneNet (Tabelini et al., 2020) and LSTR (Liu et al., 2021) based on the transformer (Vaswani et al., 2017) proposed to regress the polynomial lane curve equation with a deep network for the first time. This kind of polynomial equation cannot be used directly to deal with both extreme cases of railway tracks (near-horizontal and near-vertical zone) at the same time, which will result in unpredictable fitting errors. Thus, they cannot be adapted to the changing inclination angles of

railway tracks in the UAV images and the accuracy of the methods still needs to be improved. For the segmentation-based methods, rather than the generally mentioned semantic segmentation, lanes are represented at the instance level (Hou et al., 2019; Abualsaud et al., 2021; Ko et al., 2022). But it is inefficient to describe the lane line as a mask that needs excessive pixel-wise classification rather than presenting the line shape. Because of the discreteness of pixels that have degrees of freedom in two spatial dimensions in the predicted masks, some post-clustering strategies are widely adopted to improve the instance accuracy (Neven et al., 2018; Ko, Lee, Azam, Munir, Jeon and Pedrycz, 2022).

To overcome this problem, anchor-based detection methods were proposed, which are also state-of-the-art models for lane detection currently. For the detection of line-shaped structures, a common practice adopted in the anchor-based lane detection methods is to sample uniformly in one spatial dimension (the vertical dimension in most cases) and determine the final location in another spatial dimension (the horizontal dimension in most cases). In this way, all the predicted points have only one degree of freedom in the dimension for locating, which can incorporate the inherent prior property of line-shaped structures into deep architectures. As shown in Figure 1, the pixel width and the inclination angles of railway tracks compared to lane detection in road traffic can be varied in the images collected by UAVs under different flight heights and attitudes. Besides, the background information can also be largely different.

Concerning all these diverse challenges, this paper proposes an anchor-adaptive railway track detection network (ARTNet) to realize full-angle railway track detection with high accuracy. The study concentrates on the detection of railway tracks for the UAV aerial images, which is realized by a group of positioning anchors with horizontal or vertical directions. These anchors with different directions are closely linked to the two branches of the network. The dimension of the customized feature maps in the network is completely decided by these anchors. The predicted locations of railway tracks are determined by the activated cells that are discretely distributed in each anchor. The main contributions and novelties of this paper are presented as follows:

(1) Concerning the problem that there is no methodology to detect railway tracks of arbitrary inclination angles, this paper proposes an anchor-adaptive dual-branch railway track detection network based on projection length discrimination, which can realize full-angle railway track detection;

(2) Aiming at the potential imbalance of the two branches in the proposed framework due to the uneven distribution of the angles of railway tracks in the dataset, a balanced transpose co-training strategy is proposed to coordinately train the whole dual-branch network and improve the prediction effect;

(3) To take full advantage of the structural feature of the proposed dual-branch framework, an extra customized transposed consistency loss is proposed to efficiently guide the training of the network and improve the detection accuracy without increasing any computational complexity.

The following parts of this paper are organized as follows: Section 2 mainly introduces the proposed deep network architecture and its corresponding co-training strategy, along with the design of the loss functions used in the training process; Section 3 makes a series of experiments and gives sufficient analysis and discussion to verify the effectiveness of the proposed formulation; Section 4 presents the conclusion of the whole paper.

2 METHODOLOGY

In this section, the existing anchor-based methods and their drawbacks are introduced first. Then this paper proposes an anchor-adaptive rail track representation method to adapt to the diversity of inclination angles of railway tracks from the perspective of UAVs. Based on that, this paper presents the anchor-adaptive dual-branch architecture dubbed ARTNet for railway track detection towards UAV aerial images. To train both branches coordinately and effectively, a balanced transpose co-training strategy is proposed. Moreover, an effective integrated loss function is designed to enhance the accuracy of the network without increasing computation complexity. For clarity, Table 1 is first presented to give descriptions of some main variables and abbreviations used in this paper.



FIGURE 2. Illustration of the traditional "vertical sampling and horizontal positioning" row anchor-based method (Qin et al., 2020), in which an image is gridded into many cells and an extra cell is attached to the end of each row of cells indicating no railway track exists in that row. (a) shows a normal case with this representation method. (b) shows a typical failure case under this formulation.

Currently, there exist two different types of anchor-based detection methods. The first kind (Chen et al., 2019; Li et al., 2020; Xu et al., 2020; Tabelini et al., 2021) performs the optimization process on the lane via regressing the coordinates relative to the detected line-shaped anchors (also called line proposal), which are adapted from the box anchor in the object detection architectures: Faster R-CNN (Ren et al., 2017) and Mask R-CNN (He et al., 2017). All the line-shaped virtual anchor lines are predefined. LaneATT (Tabelini, Berriel, Paixão, Badue, Souza and Oliveira-Santos, 2021) has achieved state-of-the-art performance on multiple lane detection datasets. The second kind (Pillion, 2019; Qin, Wang and Li, 2020; Yoo et al., 2020) usually first grids the image into many cells and each row of these cells are called a row-anchor. Therefore, this method predicts the line location at each row and performs row-anchor based classification of the gridded image. A normal case is presented in Figure 2(a). All the lanes are sampled once at each row anchor. The constraints in terms of continuity and the whole shape of the lane can be added to the formulation easily.

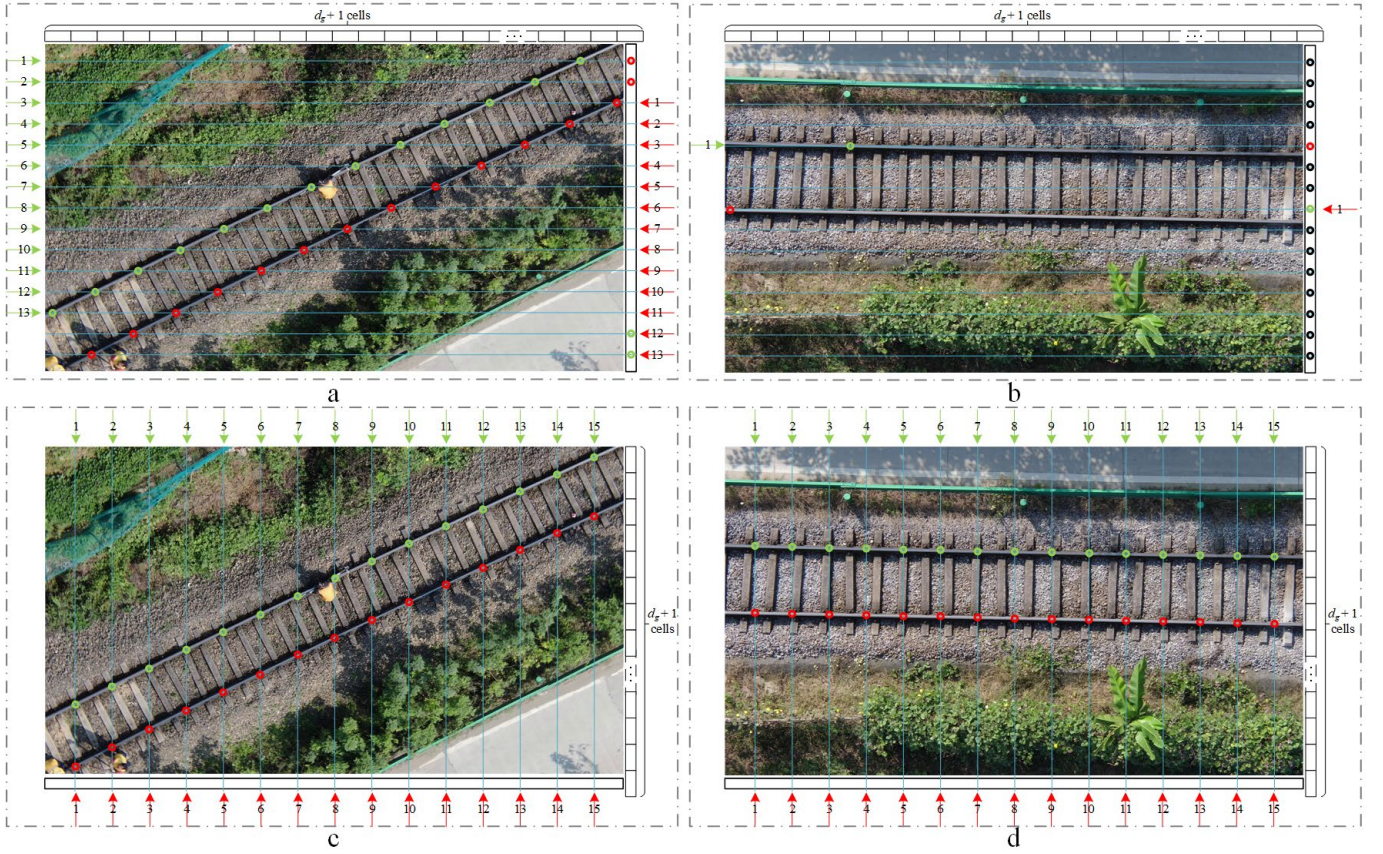


FIGURE 3. Two scenarios for anchor-adaptive railway track representation based on projection length discrimination with a sampling rate of 15 ($d_s = 15$). The first row (a, b) and the second row (c, d) show the representation of horizontal and vertical positioning anchors, respectively. The first column (a, c) is for scenario 1 and the second column (b, d) is for scenario 2. All these horizontal or vertical predefined anchors are gridded into d_g+1 separated cells (or locations). An extra gridding cell is appended to the end of each positioning anchor to denote the absence of railway tracks. Green circles and red circles indicate the location of railway tracks in the image. The black circles indicate the location where blue circles and green circles are overlapped. The maximum count on the arrow indicates the projected length in the orthogonal direction to the positioning anchors' direction.

In Figure 2, the blue and green cells indicate the location of the two tracks, respectively. The red cells indicate the location the colors blue and green are overlapped. Despite some row anchor-based models having achieved advantages in both accuracy and speed in dealing with lane detection from the perspective of cars, they still cannot address the problem of directional diversity of railway tracks from the perspective of UAVs.

In the task of lane detection, the car-mounted front camera is usually fixed and has a stable field of vision. Therefore, except for the dynamically changing traffic flow on the road, the approximate direction of lanes of the road scene as background remains unchanged. It is just because of this characteristic that almost all anchor-based lane detection methods take the "vertical evenly spaced sampling and horizontal positioning" as a natural pattern for lane representation (Li, Li, Hu and Yang, 2020; Qin, Wang and Li, 2020; Tabelini, Berriel, Paixão, Badue, Souza and Oliveira-Santos, 2021). However, the railway tracks from the perspective of UAVs usually don't match the criteria, and they always have various inclination angles in UAV aerial images. As a typical failure example shown in Figure 2(b), when the direction of the tracks is almost horizontal, a single railway track is represented by only one cell or location, which cannot be used to represent a whole railway track

obviously. This indicates the shortcomings of the traditional lane representation method.

TABLE 1. Description of variables and some abbreviations used in this paper.

Variable	Description
h	Height of the input image.
w	Width of the input image.
d_s	The number of positioning anchors.
d_g	The number of gridded cells for positioning anchors.
N	The maximum number of railway tracks in the dataset.
l_v	Mean projection length of all predicted track proposals.
l_h	Mean projection length of all predicted track proposals.
s_i	Start location of the i -th predicted railway track.
l_i	Projection length of the i -th predicted track proposal.
DBA	Dual-branch architecture.
BTCS	Balanced transpose co-training strategy.
TCL	Transposed consistency loss.
PLD	Projection length discrimination.
ARTNet	Anchor-adaptive railway track detection network.

2.1 Anchor-adaptive rail track representation

The inclination angles of railway tracks can be various in the images collected by UAVs at different flight heights and attitudes. When the angle between the direction of railway

tracks and the direction of positioning anchors becomes relatively small or even close to zero, the traditional anchor-based representation fails to represent the railway tracks effectively. Thus, like the concept of network ensemble

methods (Alam et al., 2019; Benamara et al., 2020; McCoy et al., 2022), this paper proposes an anchor-adaptive dual-branch railway track representation method for aerial images.

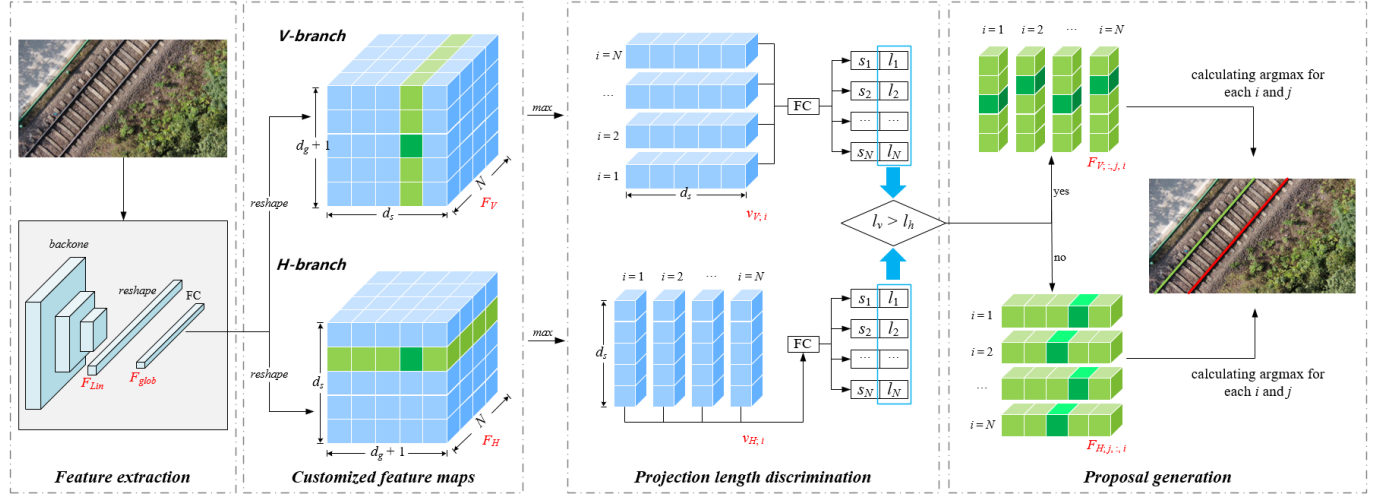


FIGURE 4. The overall architecture of the proposed ARTNet. There exist two branches in this network, i.e., H-branch and V-branch. The H-branch and V-branch correspond to the horizontal and vertical positioning anchors, respectively. The green layers in the features of the two branches show the direction of anchor selection (or classification).

A railway track is represented by a group of ordered 2D points which are uniformly sampled from the image as illustrated in Figure 3. The order pattern of these points is determined by the principle of “left to right, top to bottom”. This order needs to be consistent across data annotation, data augmentation, and target prediction. The sampling process is completed by a series of positioning lines, the directions of which can be horizontal or vertical. All these positioning lines are defined as anchors. The predefined positioning anchors are discretely and uniformly distributed in the image in the horizontal or vertical direction, the number of which is denoted as dimension d_s . Usually, the number of predefined positioning anchors is far smaller than the image size, which means: $d_s \ll h$ and $d_s \ll w$. To formulate the positions effectively, all positioning anchors are gridded into a lot of cells or locations. That means each positioning anchor is gridded into d_g+1 little cells. Then for each positioning anchor, a location (or cell) can always be found to best represent the location of the railway track in that positioning line. An extra gridding cell is appended to the end of each positioning anchor to denote the absence of railway tracks. In this way, a railway track in the image can be formulated as the set of locations of these cells in all positioning anchors, denoted as $P = \{Loc_i\}_{i=1}^{d_s}$, in which $Loc_i \in \{1, 2, \dots, d_g, d_g+1\}$. However, the direction of the positioning anchors can be horizontal or vertical. It is believed that the more sampled points there are, the closer it is to the true shape of railway tracks. Thus, given P_h and P_v denoting the representation under horizontal positioning anchors and vertical positioning anchors respectively, the final representation of the railway track can be determined by projection length discrimination:

$$P = \begin{cases} P_h, & l_h \geq l_v \\ P_v, & l_h < l_v \end{cases} \quad (1)$$

in which l_h and l_v indicate the mean value of the projected lengths of the N railway tracks in the orthogonal direction to the positioning anchors' direction respectively. The length is not the pixel length in the image but the counted number of valid positioning anchors. In Figure 3, the projection length of (a), (b), (c), and (d) are 13, 1, 15, and 15 respectively. That also means l_h and l_v for scenario 1 are 13 and 15 respectively, while l_h and l_v for scenario 2 are 1 and 15 respectively. Therefore, the results in the second row for scenarios 1 and 2 will be chosen as the final representations.

2.2 Railway track detection network

To adapt to the proposed representation method, this paper proposes a novel anchor-adaptive dual-branch railway track detection network, as presented in Figure 4. The network mainly includes four parts: feature extraction, customized feature maps, projection length discrimination, and proposal generation. An overall flow chart of these four parts is depicted in Figure 5. Firstly, the image is input to the network for feature extraction. Then these features are processed further and reshaped into customized feature maps with specific dimensions. Next, the feature maps are used to complete projection length discrimination and proposal generation, respectively. The discrimination results will be also used for the final proposal generation.

The first step of the presented ARTNet is the backbone for feature extraction, which can be achieved by any commonly used CNN like ResNet (He et al., 2016). The output of this step is a down-sampled feature map $F_{back} \in \mathbf{R}^{C \times H \times W}$ from which the features corresponding to each anchor will be generated from a max-pooling process, as described in the following part. For the sake of reducing the dimension, a 1×1 convolution is appended to produce a channel-reduced feature map $F_R \in \mathbf{R}^{C \times H \times W}$, which can bring a large reduction in computational cost. Then F_R is flattened and reshaped as a

feature vector $F_{Lin} \in \mathbf{R}^{C \cdot H \cdot W}$. Then a fully connected layer functioning to the vector F_{Lin} will generate a new global feature vector $F_{glob} \in \mathbf{R}^{d_s \cdot (d_g + 1) \cdot N}$, which can be described by

$$F_{glob} = WF_{Lin} + \mathbf{b} \quad (2)$$

in which $W \in \mathbf{R}^{d_s \cdot (d_g + 1) \cdot N \times C \cdot H \cdot W}$, $\mathbf{b} \in \mathbf{R}^{d_s \cdot (d_g + 1) \cdot N}$, and N denotes the maximum number of railway tracks in the dataset. Finally, the F_{glob} is utilized to form two branches for anchor-adaptive railway track detection, which are called H-branch and V-branch, respectively.

To cope with the anchor-based representation of railway tracks, the detection is formulated into a selecting task based on the extracted global features F_{glob} based on the positioning anchors. As mentioned above, each positioning anchor is gridded into d_g cells and for each positioning anchor, a location (or cell) can always be found to best represent the location of the railway track in that positioning line. An extra cell is appended to the end of each anchor to denote the absence of railway tracks in that positioning anchor. Hence, there are $d_g + 1$ cells in each anchor to determine the best location of a single railway track. Based on that, the F_{glob} is reshaped into F_H and F_V respectively, where $F_H \in \mathbf{R}^{d_s \times (d_g + 1) \times N}$ and $F_V \in \mathbf{R}^{(d_g + 1) \times d_s \times N}$. With F_H and F_V , the locations of the railway track and their projection lengths can be determined. All cells that are distributed in the positioning anchor correspond to certain points that are uniformly distributed in the image. These two different types of cells correspond to two different types of points with different degrees of density. But these different types of points are not good or bad because there is always an optimal way of railway track representation existing between them.

It should be noted firstly how to use F_H and F_V to determine the exact positions of the cells, i.e., P_h and P_v , for the railway tracks in the image. It is easy to recognize that all the gridded cells for all d_s anchors and N tracks are in one-to-one correspondence with the feature values in F_H and F_V . Here, F_H and F_V correspond to the horizontal and vertical positioning anchor-based railway track representation respectively. Thus, the predicted locations of the i -th track based on F_H can be determined by:

$$P_h(i, j) = \underset{k}{\operatorname{argmax}} F_{H; j, k, i} \quad (3)$$

where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, d_s$ and $k = 1, 2, \dots, d_g + 1$. The argmax function is used to find the index of the maximum value in a vector. In the same way, the predicted locations of the i -th track based on F_V can be determined by:

$$P_v(i, j) = \underset{k}{\operatorname{argmax}} F_{V; k, j, i} \quad (4)$$

where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, d_s$ and $k = 1, 2, \dots, d_g + 1$. In fact, our formulation is to choose the best positions of railway tracks for every positioning anchor with the extracted global features. Railway tracks are defined as many gridded cells on

the predefined positioning anchors, i.e., horizontal or vertical positioning anchors. Therefore, railway track detection can be determined by choosing the best cells on these predefined positioning anchors, as presented in Figure 4. The green layers in the features F_H and F_V show the direction of anchor selection (or classification), along which the predefined anchors are gridded into $d_g + 1$ separated cells. Generally speaking, our approach is composed of $(d_g + 1)$ -dimensional classifications of a dual branch network based on anchor-adaptive railway track representation.

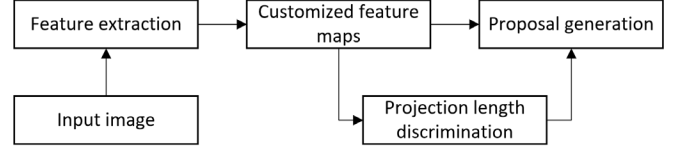


FIGURE 5. Flow chart of main parts of the proposed ARTNet.

It is also worthy to illustrate how to calculate the projection lengths of railway tracks with the extracted global features F_H and F_V , so as to fulfill the adaptive discrimination of the appropriate direction of positioning anchors. Here, it is expected to calculate the projection lengths of the railway tracks in two different directions in a regression way, while keeping a low computational cost. Inspired by the anchor-based feature pooling method (Tabelini, Berriel, Paixão, Badue, Souza and Oliveira-Santos, 2021), in which every line-shaped anchor will have its corresponding feature vector (column-vector notation) pooled from the extracted features by backbone that carries local feature information, it is supposed to extract the max-pooled feature vectors from F_H and F_V to regress the starting position s_i and the projected length l_i of the railway tracks, $i = 1, 2, \dots, N$. The vector that is max-pooled from F_H to regress s_i and l_i is obtained by:

$$v_{H; i} = \max_k F_{H; :, k, i} \quad (5)$$

where $i = 1, 2, \dots, N$ and $k = 1, 2, \dots, d_g + 1$. In the same way, the vector that is max-pooled from F_V to regress s_i and l_i is obtained by:

$$v_{V; i} = \max_k F_{V; k, :, i} \quad (6)$$

where $i = 1, 2, \dots, N$ and $k = 1, 2, \dots, d_g + 1$. When $v_{H; i}$ and $v_{V; i}$ are pooled from the two global feature maps, the fully connected layer is followed to finish the regression step. The max-pooled $v_{H; i}$ and $v_{V; i}$ contains the location information of the railway tracks and thus can be used to calculate the starting locations and the projection length values. The designed max-pooled vector can largely reduce the computation of the regression layer from $N \times d_s \times (d_g + 1)$ to $N \times d_s$, in which d_s is far smaller than d_g usually to pursue the high precision of railway track detection task. It also should be noted that the max pooling here is not the kind used for dimensionality reduction in ordinary convolutional neural networks but refers to the process denoted in (5) and (6). The predicted starting locations s_i and projection length l_i can be obtained by:

$$(s_i, l_i) = FC(v_i) \quad (7)$$

where v contains two forms, i.e., $v_{H;i}$ and $v_{V;i}$, corresponding to the two branches respectively. FC indicates the fully connected layer in the network. The regressed s_i and l_i designate a rectangular area in the image where the corresponding i -th railway track exists. The mean value of the projected lengths l_i of multiple railway tracks for the two branches can be calculated, denoted as l_h and l_v . In line with the proposed anchor-adaptive railway track representation method, the final prediction of the railway tracks $P = \{Loc_i\}_{i=1}^{d_s}$ can be obtained by formulas (1), (3), and (4).

F_H and F_V in the ARTNet are two customized feature maps. Their spatial dimensions, i.e., d_s , d_g , are completely determined by the railway track representation process. d_s is the sampling dimension for the input image and d_g is the gridding dimension for each positioning anchor, as listed in Table 1. As illustrated in Figure 4 and Figure 5, on the one hand, these two feature maps are adopted to regress the projection lengths of the railway tracks in the image; on the other hand, they are also used to generate the final railway track prediction based on projection length discrimination. In this way, the whole network can achieve anchor-adaptive railway track detection. Therefore, ARTNet is designed to cooperate with the proposed railway track representation method to realize railway track detection.

It is easy to find that the proposed algorithm is much easier than the conventional segmentation-based methods which classify at the pixel level. It turns out that our ARTNet also has a lower computational cost. Supposing that the image size is $h \times w$ and the max number of the railway tracks in a single image is N , the original segmentation-based methods need to perform classifications in $N+1$ dimension for $h \times w$ times, while ours needs only to conduct classifications in $d_g + 1$ dimension for $d_s \times N$ times. Therefore, our method needs $d_s \times N \times (d_g + 1)$ times of fundamental calculation, but the segmentation-based method will need $h \times (N+1) \times w$ times of this kind of calculation. In common, the number of predetermined positioning anchors and gridded cells is far smaller than the input image size, i.e., $d_s, d_g \ll h, w$. For example, with d_s, d_g, h, w and N being 15, 150, 512, 512, and 2 respectively, the computational cost for only the classification of our method is 4.53×10^3 while the cost of the segmentation-based method is 7.91×10^5 . The computational cost is considerably reduced which implies the superior speed performance of our proposed ARTNet.

2.3 Balanced transpose co-training strategy

As formulated, the optimal positioning anchors can always be selected to perform railway track detection by projection length discrimination, in which case the two branches of the network are designed to be trained simultaneously in a weighted way. They will be assigned different weights for optimization in the training process, where the weights are determined by the projection lengths (see Section 2.4 for details). This means that the branch corresponding to a larger

projection length contributes more to the loss function and thus receives a larger weight to be optimized.

Despite the seemingly perfect two-branch architecture capable of detecting railway tracks in an anchor-adaptive manner based on projection length discrimination, each image can only contribute more to the parameter optimization of the branch with a larger projection length. A larger projection length means a larger weight for the parameter optimization during the training process. This means that if the angular distribution of the railway tracks in the dataset is not uniform, then there is a high probability that the two branches of the trained network have different capabilities for railway track detection. In this case, one branch is trained with higher capability and accuracy, while the other has not been optimized adequately yet. This can lead to unbalanced training of the two branches of the network.

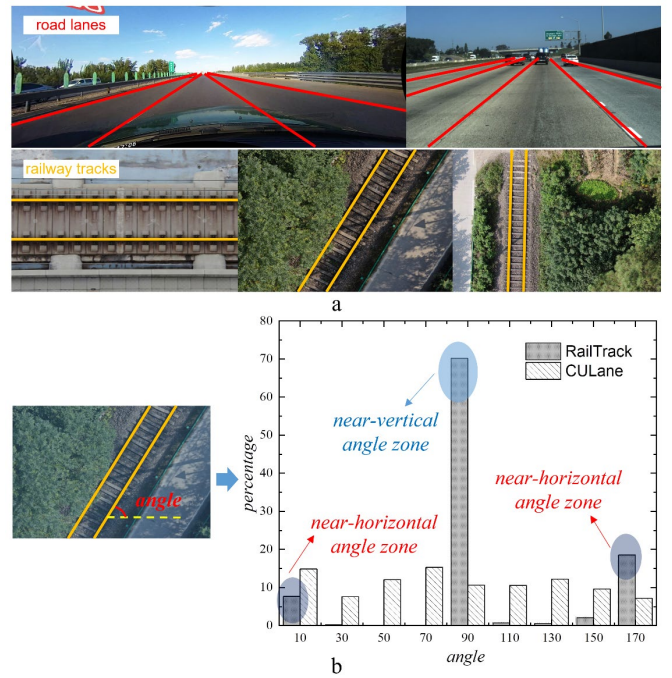


FIGURE 6. (a) Schematic diagram of angle distribution of railway track images taken by UAVs and ordinary road lane images taken by car-mounted cameras. (b) Definition for the inclination angle of railway track or road lane marker and angle distribution comparison between the constructed RailTrack and CULane.

Especially, if the projection lengths of the two kinds of representations differ greatly, in which situation the inclination angle is close to 0° , 180° , or 90° , the training of the whole network can be only concentrated to one branch. Hence, there exist many such images in our dataset that can lead to unbalanced training of the two branches because of the unbalanced distribution of the inclination angles of railway tracks. Figure 6 gives an intuitive impression of the angle distribution comparison between railway tracks and road lane markers. As shown in the bottom left figure, the angle formed by the straight line of the railway track and the horizontal rightward direction is defined as the angle of the railway track. Also, the definition is employed to define the angle of the road lane marker similarly. The angle calculation of the curved one is done by simple linear regression. As illustrated in the bottom right figure, the distribution of

inclination angles of railway tracks in the images of our dataset is statistically counted. The angles of the railway track in RailTrack are mainly distributed in the near-horizontal and near-vertical zones. Compared to CULane, the distribution of the inclination angles of the railway tracks in the constructed RailTrack is extremely unbalanced. However, the pictures of railway tracks with extreme inclination angles are usually what the UAV-based railway inspection really needs, which poses a big challenge to the detection architecture.

To deal with the potential imbalanced training problem, a balanced transpose co-training strategy is proposed to train the two branches coordinately. The transpose transformation can keep the order, i.e., left to right and top to bottom, of the sampled points of railway tracks unchanged, as shown in Figure 7. This special kind of order-preserving property is not possessed by other transformations such as rotation and mirroring. Two different examples of inclination angles of railway tracks are presented in Figure 7. For the left image in Figure 7(a), the representation of railway tracks adopts the horizontal positioning anchors. Correspondingly, such images will contribute more to the training and optimization of the H-branch. If this representation and the image itself are transposed, the horizontal positioning anchor-based representation of the railway tracks naturally becomes the vertical positioning anchor-based representation, which means that the transposed image can also be used to train the V-branch of the network with a larger weight. Similarly, for the right image in Figure 7(b), the representation of railway tracks adopts the vertical positioning anchors and such images will contribute more to the training and optimization of the V-branch. If this representation and the image itself are transposed, the vertical positioning anchor-based representation naturally becomes the horizontal positioning anchor-based representation, which means that the transposed image can be used to train the H-branch of the network with a larger weight.

It can be concluded that if the anchor-adaptive representation of railway tracks based on the projection lengths is P for an image Q that can be used to mainly train one branch of the network, then the transposed image Q^T , with the corresponding representation denoted as P^T , can be exploited to focus on training the other branch of the network. In this way, two different forms of the same image can be used to solve the problem of unbalanced training of the two branches caused by the extremely unbalanced distribution of inclination angles of the railway tracks in the dataset. Benefiting from the proposed balanced transpose co-training strategy, both branches can be effectively trained and will have consistent prediction ability in theory, which can meet the requirements of the proposed anchor-adaptive dual-branch architecture based on projection length discrimination.

Suppose the horizontal anchor-based representation of the image Q as $P = P_h = \{Loc_{H,i}\}_{i=1}^{d_s}$ and the vertical anchor-based representation of Q^T as $P^T = P_v^T = \{Loc_{V,j}^T\}_{j=1}^{d_s}$, then for $i = j$ the following formula holds:

$$Loc_{H,i} = Loc_{V,j}^T \quad (8)$$

where $i, j \in \{1, 2, \dots, d_s\}$ as shown in Figure 7(a). Similarly, the following formula holds as well:

$$Loc_{V,i} = Loc_{H,j}^T \quad (9)$$

where $i, j \in \{1, 2, \dots, d_s\}$ as shown in Figure 7(b).

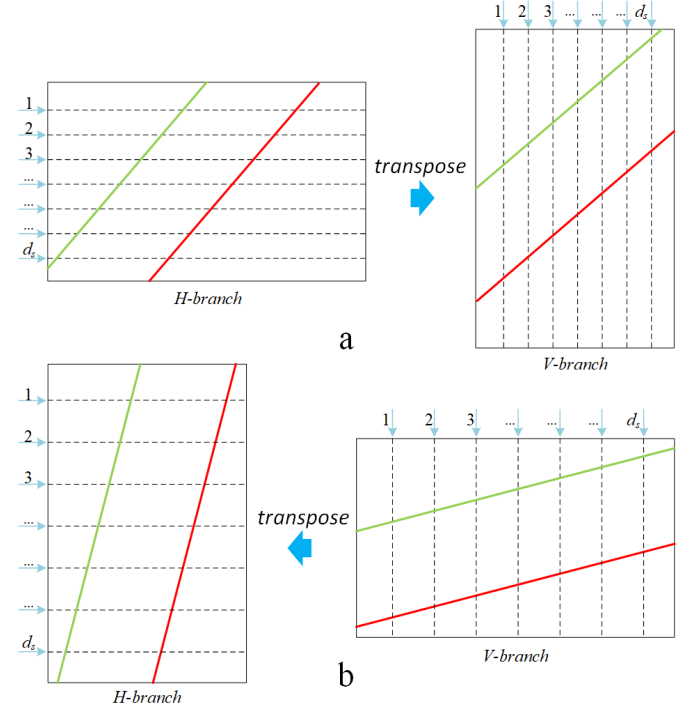


FIGURE 7. The illustration of the balanced transpose co-training strategy and its order-preserving properties for the sampled point groups.

2.4 Design of loss functions

For efficient and balanced training of the two branches and the backbone ahead of them, an integrated loss function is presented to accomplish the training and the optimization of the architecture. It mainly includes three parts: the weighted classification loss L_{cls} based on projection length discrimination, the smooth L1 regression loss L_{reg} , and the proposed transpose consistency loss L_{TC} . Since there are two branches in the network, they can be used for track detection based on horizontal anchors and vertical anchors respectively. However, for a specific image, it is supposed that these two branches have different weights to be trained and optimized because of the different projection lengths correspondingly. The weights for these two branches to classify can be calculated by the softmax function:

$$w_h = e^{l_h} / (e^{l_h} + e^{l_v}), \quad w_v = e^{l_v} / (e^{l_h} + e^{l_v}) \quad (10)$$

where e is the natural constant. As listed in Table 1, l_v is the mean projection length of all predicted track proposals and l_h is the mean projection length of all predicted track proposals. w_h and w_v are computed to weigh the losses of the two classification branches. Thus, the final weighted classification loss can be formulated by:

$$L_{cls} = w_h L_{cls_h} + w_v L_{cls_v} \quad (11)$$

where L_{clsh} and L_{clsv} are both the Focal Loss (Lin et al., 2020) to constrain the classification of H-branch and V-branch in the network, respectively. L_{reg} adopts the Smooth L1 loss function to achieve the regression of the starting positions and projection lengths in the two branches, which can be calculated as follows:

$$L_{reg} = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)| \quad (12)$$

in which x_i represents the vector $v_{H,i}$ or $v_{V,i}$. y_i indicates the regressed values including s_i and l_i .

Now supposing the prediction of the original image Q and the transposed image Q^T , i.e., the j -th sampled point of the i -th railway track in the image, are denoted with $Loc_{i,j}$ and $Loc_{i,j}^T$, respectively, then a novel loss function is easily obtained to restrict the predicted results of the two branches for Q and Q^T according to the relative correspondence implied in Section 2.3. Thus along with the proposed balanced transpose co-training strategy, a transposed consistency loss function L_{TC} is presented which can be calculated by:

$$L_{TC} = \frac{1}{Nd_s} \sum_{i=1}^N \sum_{j=1}^{d_s} (\|Loc_{H,i,j} - Loc_{V,i,j}^T\| + \|Loc_{V,i,j} - Loc_{H,i,j}^T\|) \quad (13)$$

Then, an integrated structural loss function to train the proposed dual-branch network is presented as follows:

$$L = \alpha L_{cls} + \beta L_{reg} + \gamma L_{TC} \quad (14)$$

in which α , β and γ are loss coefficients.

3 EXPERIMENTS

In this section, a RailTrack dataset is constructed to perform UAV-based railway track detection, and the general setup of the relevant parameters is presented. Then the evaluation metrics adopted in this paper are introduced. The ablation study on the proposed dual-branch architecture, the balanced transpose co-training strategy, and the transpose consistency loss of ARTNet are conducted to validate the performance of these designs. This paper has also made comprehensive comparisons with other popular algorithms and some visual examples are presented.

3.1 Dataset and parameter setting

This paper has built a railway track detection dataset named RailTrack, in which all images are collected from some ordinary railway lines and Beijing-Shanghai high-speed railway line. All the images are taken by the Zenmuse H20T aerial camera, which is mounted on the DJI Matrice 30 RTK flying platform. The H20T is equipped with a 20-megapixel zoom camera that can achieve $23\times$ hybrid optical zoom, which enables it to see the railway tracks clearly at high altitudes. The relative distance between the UAV and the railway plane is generally set to 50-80m and the lateral distance is usually set to 10-20m to ensure the safety of the railway's normal operation. All images in the RailTrack are taken with this distance setting. The RailTrack dataset is divided into the training part (893) and test part (223)

including various kinds of images in which the railway tracks appear at different inclination angles and scales, as shown in Figure 1. All the images in the dataset have a resolution of 1920×1080 . The background information and the pixel width of the railway tracks in the dataset are also abundant and changeable for the sake of enhancing the generalization ability of the trained models. Figure 8 presents some photos of the image collection process with UAVs. The Matrice 300 RTK equipped with Zenmuse H20T in the left image is both industry-level equipments produced by DJI company.

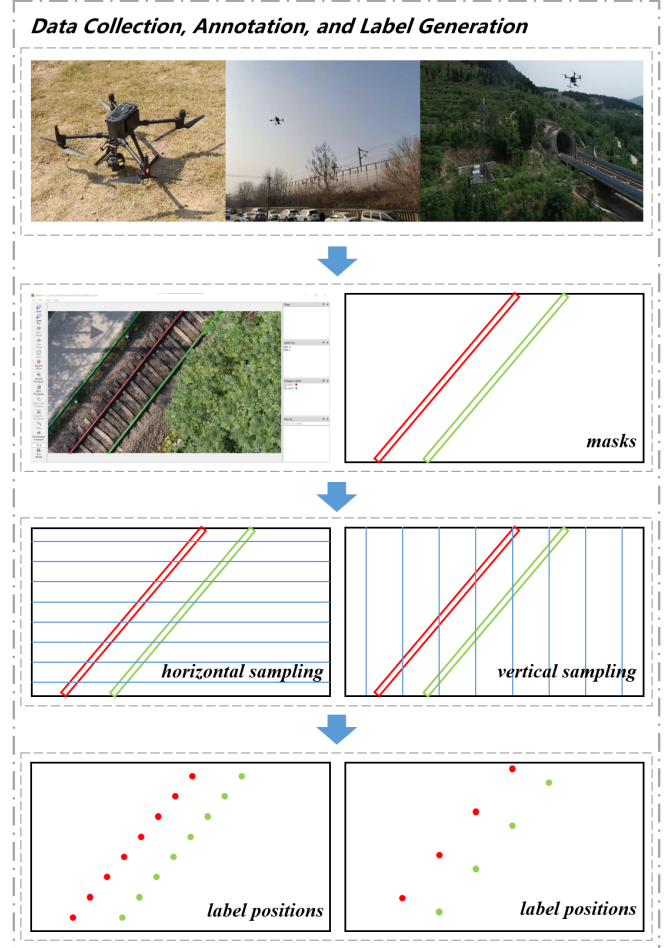


FIGURE 8. Illustration of data collection, data annotation, and label position generation process.

In this work, the formulated ARTNet and other comparative methods are trained with the same batch size (16) setting and input size (512×512) setting. Empirically, stochastic gradient descent is executed by the same Adam optimizer. And in the following ablation study section, all the trained networks are with a d_s of 15 and a d_g of 150. The deep modules in this work are all implemented with the deep learning framework PyTorch.

As shown in Figure 8 some schematic working photos for data collection with UAVs are presented. Then the data annotation and label generation process are introduced. There are three steps in the process: (1) annotating to make railway track masks for the original images; (2) sampling the masks with horizontal and vertical anchors; (3) generating the final label positions. The data annotation adopts the software LabelMe to produce masks of the original images by drawing

polygons on them. The generated label positions are used to calculate the proposed integrated structural loss and perform prediction evaluation. During the data annotating process, the coupled railway tracks are labeled in the order of "left-to-right, top-to-bottom".

3.2 Evaluation metrics

The most used metric in lane detection is the F1 score used in the CULane. Each lane marker is considered a line with a width of 30 pixels in the evaluation stage in CULane. Intersection over union (IoU) between the prediction and corresponding ground truth is calculated to determine the quality of the prediction. The prediction with IoU larger than θ is treated as a true positive (TP), and the one with IoU smaller than θ is treated as a false positive (FP). The ones that exist but are not detected are counted as false negatives (FN). the IoU threshold θ is predefined and usually taken as 0.5 in CULane. In this paper, an F1-related metric named mF1 (Zheng et al., 2022) is adopted to better compare the performance of different algorithms. The mF1 is defined as:

$$mF1 = (F1@30 + F1@50 + F1@75) / 3 \quad (15)$$

where $F1@30$, $F1@50$, and $F1@75$ are F1 metrics when IoU thresholds are 0.3, 0.5, and 0.75 respectively.

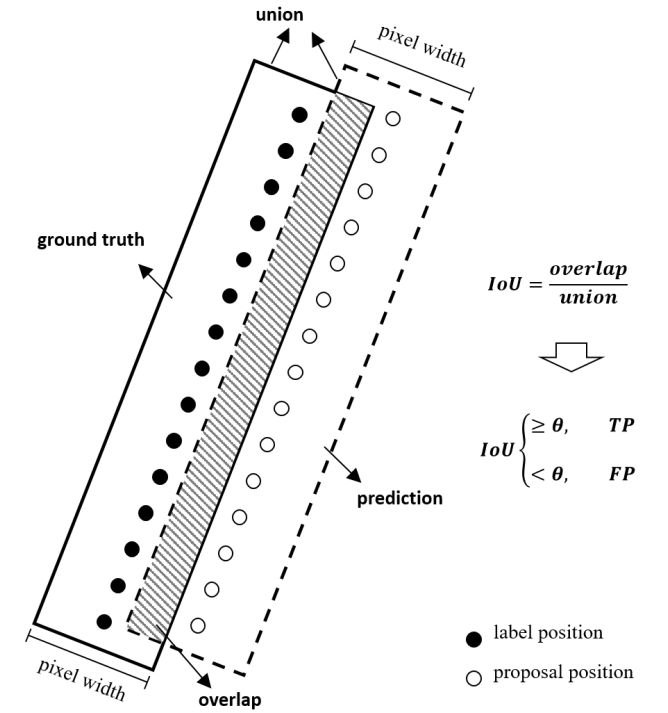


FIGURE 9. The IoU calculation between the proposal positions and their corresponding label positions. The proposal with $\text{IoU} \geq \theta$ is treated as a true positive (TP) and the one with $\text{IoU} < \theta$ is treated as a false positive (FP).

It should be noted that the output of the anchor-based method is usually not an array of pixels indicating the area of railway tracks or lanes but a group of proposal points. Figure 9 has illustrated the IoU calculation between the proposal positions and their corresponding label positions. These points in this paper can be mapped by the activated cells

existing in the positioning anchors. Each activated cell corresponds to a certain point in the input image. All these points constitute the final railway track proposal. Then a prediction rectangle area centered on these proposal points will be generated. In the same way, a ground truth rectangle can be generated as well. Both rectangles have the same pixel width. The IoU between the prediction rectangle and its corresponding ground truth rectangle will be calculated.

But the pixel widths of railway tracks can be various in the UAV remote sensing images as illustrated in Figure 1. Therefore, to balance the validity and effectiveness of IoU calculation and the distribution of the pixel widths of the railway tracks in the constructed dataset, statistics on the pixel widths of the railway tracks in the dataset are counted. The widths of the tracks are mostly concentrated in 10-15 pixels, followed by 15-20 and 25-30 pixels. The calculated mean value, 1/4 quantile, 2/4 quantile, and 3/4 quantile are 20, 12, 15, and 28 respectively. Since the 3/4 quantile (28 pixels) can cover most cases of the railway tracks in our dataset, it is more suitable for the final evaluation of the generated proposals. Thus, this paper takes the 3/4 quantile to be the final pixel width to calculate the IoU metric for the predicted railway tracks and corresponding ground truths.

3.3 Ablation study

The effectiveness of the proposed three main designs in our formulation, i.e., dual-branch architecture, the balanced transpose co-training strategy, and the transpose consistency loss, have been investigated and verified. It should be noted that the balanced transpose co-training strategy can only be adopted when the dual-branch architecture design is used. The transposed consistency loss can only be adopted when the balanced transpose co-training strategy design is used. This paper takes the model proposed by (Qin, Wang and Li, 2020) as the baseline. As shown in the top three rows in Table 2, three baseline models backbone by resnet18, resnet34, and resnet50 are first evaluated. All three models are randomly initialized and trained for 500 epochs and the results are far from satisfactory. The three baseline models have only one branch to detect the railway tracks in the image and no transposed image is used. These results illustrate that it is the architecture itself rather than the backbone network for feature extraction that limits the final capability of the baseline models. Noted that all the comparative models presented in Table 2 are trained for the same 500 epochs.

Ablation on the dual-branch architecture. The architecture is designed to fulfill the adaptive representation and prediction of railway tracks. This kind of design makes the network into two branches: H-branch and V-branch, which positions railway tracks in horizontal or vertical direction respectively. When performing forward inference, the network adaptively selects a branch to predict the railway tracks according to their projected lengths in horizontal and vertical directions. With this adaptive design applied to the baseline model (rows 1-3 of Table 2) as shown in row 4 of Table 2, the F1-score-related values achieve an overall

improvement. Especially the mF1 value increased from 12.51 to 51.35, which demonstrates the huge advantage and effectiveness of this dual-branch architecture design. In this part, only the original images are involved in this experiment without using any transposed images.

In addition, the single-branch baseline trained with the

simple transposed augmentation design is also evaluated, in which the original images and their transposed ones are both fed into the architecture to detect railway tracks by the final projection length discrimination. As given in row 7 of Table 2, the result indicates the effectiveness and advancement of the dual branch network.

TABLE 2. The ablation study results for the proposed three designs: the dual-branch architecture (DBA), the balanced transpose co-training strategy (BTCS), and the transposed consistency loss (TCL). A simple transposed augmentation (STA) design applied to the single-branch baseline is also evaluated. The projection length discrimination (PLD) is used accordingly.

No.	model	backbone	initialize	DBA	BTCS	TCL	PLD	F1@30	F1@50	F1@75	mF1
1	baseline	resnet18	Rand.	×	×	×	×	19.51	13.72	4.31	12.51
2	baseline	resnet34	Rand.	×	×	×	×	19.27	14.34	3.74	12.45
3	baseline	resnet50	Rand.	×	×	×	×	19.38	14.74	4.14	12.75
4	+DBA	resnet18	Rand.	✓	×	×	✓	69.84	57.90	26.31	51.35
5	+DBA+BTCS	resnet18	Rand.	✓	✓	×	✓	72.22	56.80	50.68	59.90
6	+DBA+BTCS+TCL	resnet18	Para.5	✓	✓	✓	✓	82.09	72.51	46.66	67.09
7	+STA	resnet18	Rand.	×	×	×	✓	75.90	63.44	31.11	56.81

Ablation on the balanced transpose co-training strategy. The strategy is designed to fulfill the balanced co-training of the two branches of ARTNet. As mentioned above, the two branches are trained based on projection length discrimination of the railway tracks. The parameters of the branch corresponding to a larger projected length of the railway tracks can be updated with a larger weight. This means the unbalanced distribution of the inclination angles of railway tracks in the dataset can lead to quite different and unbalanced prediction capabilities of the two branches. To deal with this unbalanced training problem, this paper proposes a balanced transpose co-training strategy to train the dual-branch network, in which way both branches of the network can be efficiently and adequately trained regardless of the inclination angle distribution of the dataset. As illustrated in rows 4 and 5 of Table 2, the F1-measure-related values achieve an overall improvement when the balanced transpose co-training strategy is applied to the basic dual-branch network. Especially, the mF1 value increased from 51.35 to 59.90, which proves the huge superiority of the proposed design of the balanced transpose co-training strategy. As required by the training strategy, the transposed images are adopted to train the two branches of the model along with the original images in this part.

Ablation on the transposed consistency loss. The loss function is also designed to guide the training of the proposed dual-branch network as part of the loss function. It is precisely because of the adoption of dual-branch architecture and balanced transpose co-training strategy that there is a certain relevance between the prediction results of two branches for the original and transposed form of the same input image. Based on this relevance, the transposed consistency loss is designed to obtain better consistency in the predictions of the two branches. As shown in rows 5 and 6 of Table 2, the F1-score-related values also achieve an overall improvement when the loss is employed. Especially, the mF1 value increased from 59.90 to 67.09, which proves the great effectiveness of the proposed design of transposed consistency loss. Here the parameters of the network are initialized with that of the trained network in row 5 of Table 2

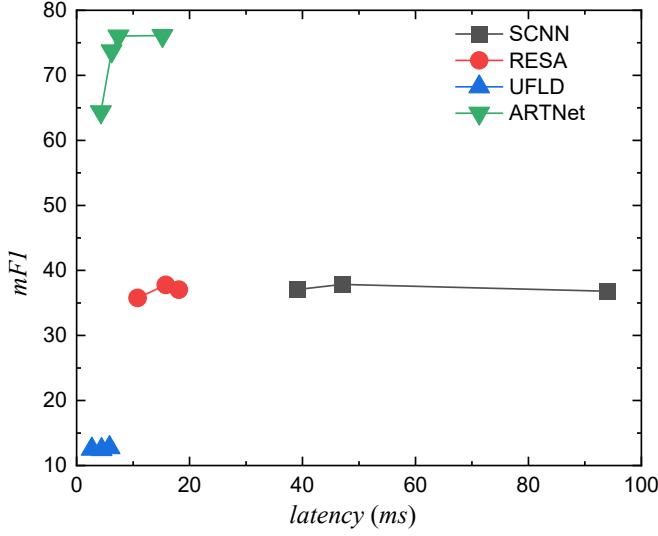
because this loss is presented in an unsupervised form. In the case of employing the transposed consistency loss, if the network is randomly initialized, the network training will likely fail to converge. α , β , and γ are all set to be 1.0 in this experiment.

TABLE 3. The forward inference speed evaluation on the dual-branch architecture (DBA) design. “3090” means NVIDIA GeForce RTX 3090 card and 2060 means NVIDIA GeForce RTX 2060 card. The frames per second (FPS) values are reported below.

backbone	baseline		+DBA	
	3090	2060	3090	2060
resnet18	372	225	233	140
resnet34	226	127	161	94
resnet50	171	72	135	61
resnet101	145	43	66	37

Ablation on the inference speed. The effects of the three designs on the inference speed are also investigated. Theoretically, the balanced transpose co-training strategy is just a kind of training strategy that will not lead to structural changes to the architecture. This means the employment of the balanced transpose co-training strategy will not slow down the inference speed of the network. In the same way, as part of the loss function, the transposed consistency loss is also a design that is only used and calculated in the training process rather than the forward inference process, which will also not make structural changes to the architecture. Thus, the transposed consistency loss will not affect the speed of the network, either. The design of dual-branch architecture changes the network from a single-branch form to a dual-branch form, which can obviously lead to a decrease in the network reasoning speed. The relevant results are shown in Table 3. The network with dual-branch architecture has a certain degree of decrease in inference speed compared with the original baseline model. However, as the backbone goes deeper, this trend becomes less pronounced. It is also illustrated that the proposed ARTNet backbone with resnet18 and employed with all three designs, i.e., the dual-

branch architecture, the balanced transpose co-training strategy, and the transposed consistency loss, can achieve an



inference speed of up to 140 FPS. The inference speed of the proposed ARTNet is fast enough for practice.

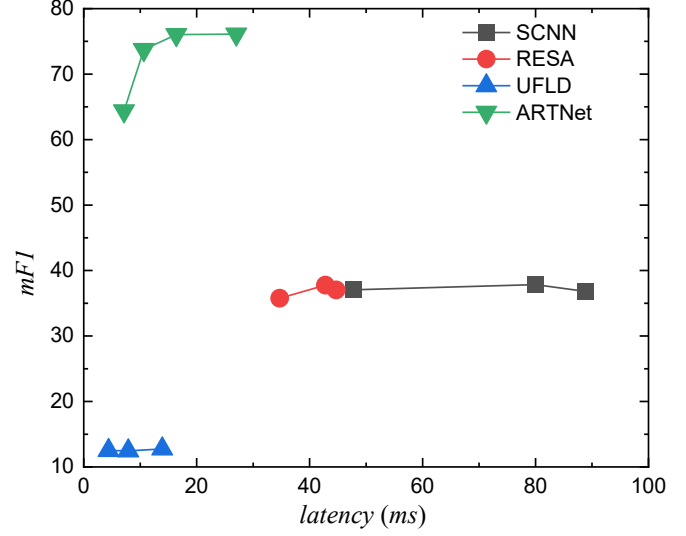


FIGURE 10. Latency vs. mF1 score of different models on both 3090 and 2060 GPU cards.

3.4 Comprehensive comparison

To demonstrate the effectiveness of the proposed architecture, comprehensive comparisons with popular lane detection methods, i.e., SCNN (Pan, Shi, Luo, Wang and Tang, 2018), RESA (Zheng et al., 2021), and UFLD (Qin, Wang and Li, 2020), have been conducted. The codes for three algorithms can be found at <https://github.com/Turoad/lanedet>. All these models are trained sufficiently with the input size of 512×512 . These models are trained multiple times and the best-performed models are adopted for comparison.

TABLE 4. Comprehensive comparisons between some popular algorithms and the proposed ARTNet.

model	backbone	F1@30	F1@50	F1@75	mF1
SCNN	resnet50	40.59	39.85	30.71	37.05
SCNN	resnet101	41.32	39.85	30.71	36.81
SCNN	vgg16	48.52	46.05	18.98	37.85
RESA	resnet18	40.51	40.15	26.64	35.77
RESA	resnet34	41.17	40.44	31.69	37.77
RESA	resnet50	40.80	40.43	29.87	37.03
UFLD	resnet18	19.51	13.72	4.31	12.51
UFLD	resnet34	19.27	14.34	3.74	12.45
UFLD	resnet50	19.38	14.74	4.14	12.75
ARTNet	resnet18	82.99	73.24	46.26	67.50
ARTNet	resnet34	86.85	81.86	58.73	75.81
ARTNet	resnet50	89.12	82.09	56.92	76.04
ARTNet	resnet101	88.44	81.86	58.05	76.12

As presented in Table 4, the UFLD is exactly the baseline model mentioned above in the ablation study section. As can be seen, the UFLD models perform badly. Its capability is not limited by the backbone which is used for feature extraction but by the architecture itself. As for the series of SCNN and RESA, both have a large gap from our expected level. It is analyzed that the performance of SCNN and RESA are affected by the existence of the near-horizontal railway tracks in the images, which can be proved by some visual examples

presented in rows 3 and 4 in Figure 11. As can also be concluded, it is not the backbone that limits the capability of SCNN and RESA but the architectures themselves.

TABLE 5. Evaluation index for different inclination angular ranges for the test set. “N-horizontal” represents near-horizontal rail tracks and is defined in the range of $0 \sim 30^\circ$ and $150 \sim 180^\circ$. “N-vertical” represents near-vertical rail tracks and is defined in the range of $60 \sim 120^\circ$.

ranges	N-horizontal	N-vertical	others	in total
TP	247	82	61	390
all	290	97	71	458
percentage	85.2%	84.5%	85.9%	85.2%

Adapted from the baseline model UFLD, the proposed ARTNet can deal with full-angle railway track detection of various UAV aerial images. As illustrated in Table 3, the proposed ARTNet backbone by resnet18 can obtain 67.50 of mF1 metric, far outperforming the comparative SCNN, RESA, and UFLD model series. When the backbone is changed to a larger one, i.e., resnet34, resnet50, and resnet101, the mF1 metric is promoted to 75.81, 76.04, and 76.12 respectively, which implies the superior performance and effectiveness of the proposed dual-branch architecture. The proposed ARTNet can efficiently address the problem that the original UFLD confronts in the transfer from lane detection task to railway track detection task toward UAV aerial images.

To compare the forward inference speed of the proposed ARTNet and other models, a series of comparative experiments on an NVIDIA Geforce RTX 3090 and 2060 card respectively with input size 512×512 . Combining the F1 scores, the computational speed results are reported. As shown in Figure 10, the latency vs. mF1 curves of different algorithms can be roughly divided into three different levels. The low-level models cannot apply to the constructed RailTrack dataset very well, including UFLD. The models in

this area are in fact not suitable for the detection of railway tracks in RailTrack and are even not suitable for the railway track detection of UAV-bases aerial images of railway scenarios. The middle-level models can recognize quite a number of railway tracks but are still limited to the detection of non-near-horizontal railway tracks, including SCNN and

RESA. The proposed ARTNet is distributed in the high-level area and outperforms all other models in terms of both latency and accuracy. More specifically, ARTNet can achieve an FPS of up to 233 for forwarding inference in the NVIDIA Geforce RTX 3090 platform compared to 26 of SCNN and 93 of RESA.

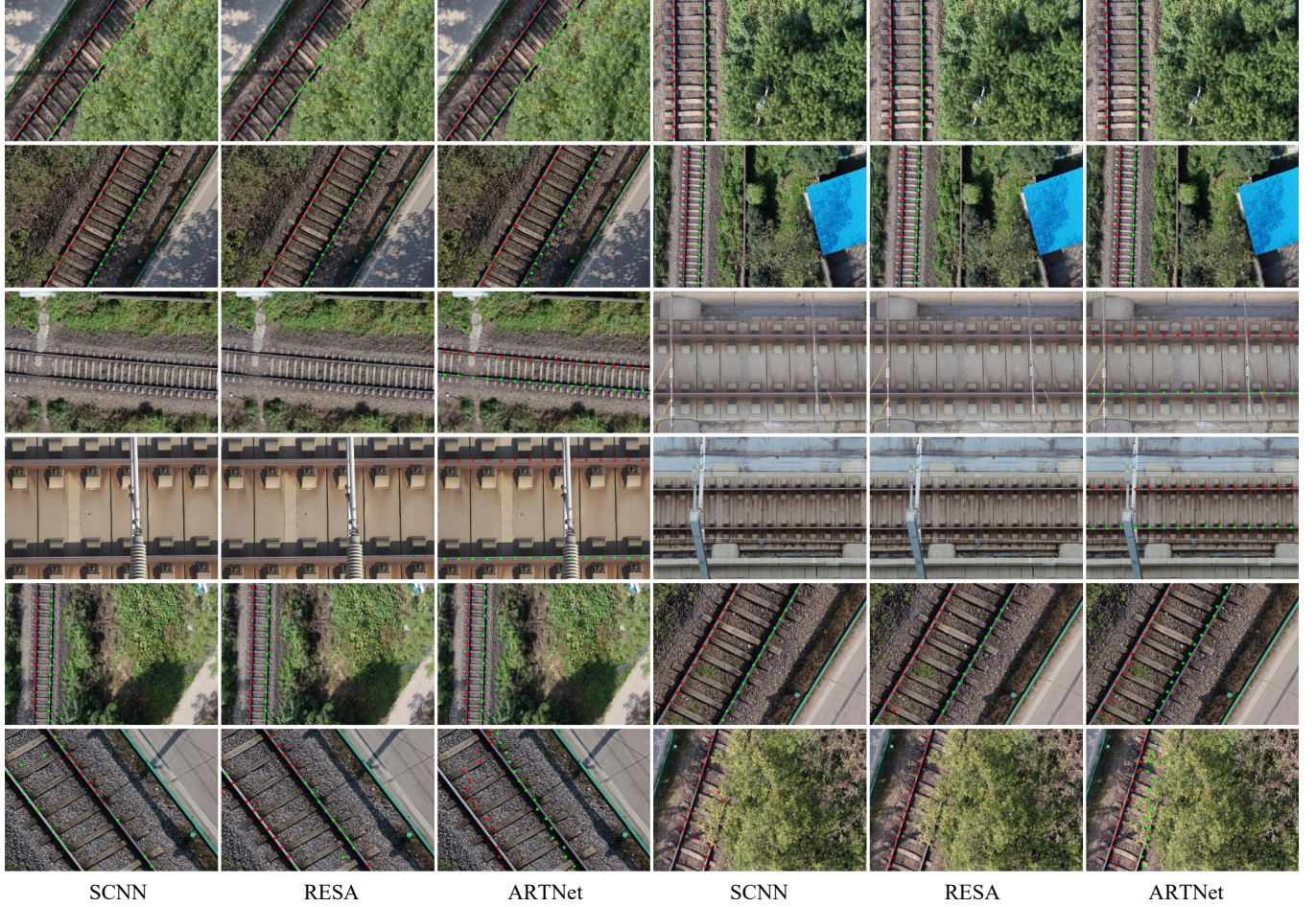


FIGURE 11. Visual examples of our formulation ARTNet compared to SCNN and RESA.

It is analyzed that the dual-branch network and anchor-adaptive railway representation proposed in this paper can deal with the railway tracks in all angular ranges including both near-horizontal and near-vertical ones and this has also been verified in our experiments. ARTNet can always select an optimal direction to generate railway track proposals by projection length discrimination. But other comparative algorithms cannot recognize the near-horizontal railway tracks under the traditional representation formulation. More evidence of visual examples can be found in the next section. Although the baseline model UFLD is faster than the adapted ARTNet, its accuracy results are far from satisfactory. In addition, some other models like LaneATT and ENet-SAD (Hou, Ma, Liu and Loy, 2019) are also included in the comparison but don't obtain satisfactory results either. The experiments illustrate the superiorities of the ARTNet on railway track detection.

Further, the predictions and their angular distributions of the 458 railway tracks in the test set of RailTrack are counted, as shown in Table 5. As can be concluded, the proposed ARTNet can deal with railway tracks distributed in various

inclination angle ranges including those with extreme angles.

3.5 Visual quality of railway track detection

Several visual examples of the proposed ARTNet and the comparative algorithms, i.e., SCNN and RESA are presented in Figure 11, in which SCNN, RESA, and ARTNet are backboneed with resnet50, resnet34, and resnet34 respectively. In general, our proposed ARTNet can effectively recognize the railway tracks of various angles, including those with near-horizontal and near-vertical inclination angles. For the examples in rows 1 and 2 and the left example of row 5, all models can recognize the two railway tracks very well despite some missed points for SCNN and RESA.

Especially, the SCNN and RESA models are completely unable to detect the railway tracks with near-horizontal angles, as shown in rows 3 and 4 of Figure 11. As expected, this is a common shortcoming of the existing methods. In contrast, ARTNet can still detect this kind of railway track with high performance. Our approach can deal with both kinds of extreme cases of railway tracks, i.e., near-horizontal and near-vertical ones at the same time as well as the ones with ordinary inclination angles. The inherent properties of

the designed adaptive anchors and the mechanism of the projection length discrimination enable the architecture to detect railway tracks of near-horizontal and near-vertical angular ranges simultaneously. This is the exact key intention of designing this adaptive full-angle detection network. Some examples with trackside poles and electric cables are also included in rows 3 and 4 of Figure 11. As can be seen, the detection results will be hardly influenced by these poles or cables. Although they are also line-shaped objects, they can hardly cause negative effects on the detection of the railway tracks in the image. These visual results have demonstrated the high performance and validity of our formulation.

In the right example of row 5 and the left example of row 6, the detection results of the two railway tracks are cross-mixed which can be caused by the unbalance of the constructed dataset. But the proposed ARTNet can alleviate this problem to a large extent due to its global feature design compared to the other two comparative networks. The bottom left example shows an example with relatively low accuracy. In this case, the predicted points for the left track deviate to the right track. It is analyzed that this type of image has certain optical distortion and is present in the dataset in small numbers,

which can possibly result in such poor prediction performance. But as can be seen, both other two models generate worse predictions with more chaotic point distributions, which indicates the superiority of ARTNet. For the bottom right example in which one of the railway tracks are severely occluded by the trees, both SCNN and RESA can detect only one railway track. Compared to them, the proposed ARTNet can recognize the existence of both two railway tracks with higher performance, occluded or not. ARTNet integrates certain fully connected layers to produce the customized global feature maps F_H and F_V . This makes both feature maps have a global receptive field on the input image, which means they can extract more useful features to discriminate the existence of railway tracks.

It is also worth noting that the proposed model is not only suitable for the detection of long straight railway tracks, but also all types of railway tracks including curves. In our formulation, each railway track is represented as a series of discrete points with sampling dimension d_s . Except for straight line-shaped railway tracks, these d_s points can also represent railway tracks in any shape. More evidence can be found in the next section.



FIGURE 12. Visual examples of ARTNet on the test set of the datasets NormalRail and Rail4Track. The left three columns are cases from the NormalRail dataset and the right three columns are cases from the Rail4Track dataset

3.6 Experiments on the generalization ability

Theoretically, the proposed ARTNet can be used not only for railway track detection in railway scenes but also for line-shaped infrastructure detection in other scenes. Thus in this section, two more datasets named Rail4Track and NormalRail are constructed to evaluate the generalization ability of the proposed detection architecture.

TABLE 6. The size of the training, validation, and test sets for the datasets Rail4Track and NormalRail.

dataset	training	validation	test	in total
Rail4Track	1290	270	290	1850
NormalRail	3500	750	800	5050

The images in Rail4Track and NormalRail are captured from the UAV-mounted camera view and the front camera view of the train respectively. Both involved up to four railway tracks (i.e., $N = 4$). As listed in Table 6, the two

datasets have 1850 and 5050 images respectively, all of which have a resolution of 1920×1080 . To improve the generalization ability of the network, the transformations of random brightness and random motion blurring are included in Rail4Track and random snow, random rain, and random fog are involved in the images of NormalRail. The splitting details of these two datasets have been presented in Table 6. All the models in this section are trained for the same 200 epochs. The network weights that perform best on the validation set during the training process will be used as the final model for evaluation and testing.

The network setting for two different datasets is illustrated in Table 7. The resnet34 is adopted to be the backbone of these two cases. Other parameter settings are the same as the setting mentioned in Section 3.1. Evaluation results on these two datasets are shown in Table 7. The proposed ARTNet can achieve an mF1 of 67.13 and 63.88 on the two different datasets respectively, which proves its capability in detecting

such line-shaped infrastructures. Some visual examples of various experimental scenarios including curved segments, rail switch segments (or Y-shaped segments), and multiple tracks are presented in Figure 12. Some examples of augmented rain, snow, fog, and motion blur are also involved. The left three columns of Figure 12 are cases from the NormalRail dataset. Both the straight ones (row 1) and the curved ones (row 2) can be recognized with high accuracy. The three well-performing examples of row 3 show the ones with augmented rain, snow, and fog respectively. The right three columns of Figure 12 are cases from the Rail4Track dataset. The first and the second row shows the scenarios for multiple tracks and rail switch (Y-shaped segment) respectively. Examples in the third row are the augmented ones with low brightness, high brightness, and motion blurring. These results have demonstrated the ability of the architecture to be applied in different railway scenarios.

TABLE 7. Evaluation results of ARTNet on the Rail4Track and NormalRail

dataset	d_s	d_g	F1@30	F1@50	F1@75	mF1
Rail4Track	15	150	86.70	74.70	40.00	67.13
NormalRail	25	250	87.00	74.69	29.94	63.88

Experiments in this section illustrate the proposed ARTNet has good generalization ability on the task of railway track detection. ARTNet provides a general railway track detection architecture for various UAV-based inspection scenarios, which has a wide range of application prospects.

4 CONCLUSION

This paper proposes a novel anchor-adaptive dual-branch architecture based on projection length discrimination to realize full-angle railway track detection for UAV-based railway aerial images. Along with the deep architecture, a balanced transpose co-training training strategy and an extra customized transposed consistency loss are proposed to guide the training of the two branches of the network in a more efficient and balanced way without increasing any computational complexity. Our approach can deal with the extreme cases of railway tracks with both near-horizontal and near-vertical inclination angles at the same time as well as the ones with other ordinary inclination angles. Experiments have demonstrated the huge superiorities of the proposed ARTNet in terms of both accuracy and inference speed, which can achieve an mF1 of 76.12 and an FPS of 66, outperforming all other comparative algorithms. The most lightweight version of the network can infer at a speed of 233 FPS. In the future, more experiments on some sophisticated learning algorithms such as neural dynamic classification, dynamic ensemble learning, and fast finite element machines will be conducted to further improve the detection accuracy and reasoning efficiency. More investigations will be performed to enhance the practical availability of the proposed method and the possibility to deploy the algorithm on UAV onboard computing devices.

REFERENCES

- Abualsaud, H., Liu, S., Lu, D. B., Situ, K., Rangesh, A. & Trivedi, M. M. (2021), Laneaf: Robust Multi-Lane Detection with Affinity Fields, *IEEE Robotics and Automation Letters*, **6**(4), 7477-7484.
- Alam, K. M. R., Siddique, N. & Adeli, H. (2019), A Dynamic Ensemble Learning Algorithm for Neural Networks, *Neural Computing and Applications*, **32**(12), 8675-8690.
- Bang, S., Hong, Y. & Kim, H. (2022), Proactive Proximity Monitoring with Instance Segmentation and Unmanned Aerial Vehicle - Acquired Video - Frame Prediction, *Computer-Aided Civil and Infrastructure Engineering*, **36**(6), 800-816.
- Behrendt, K. & Soussan, R. (2019), Unsupervised Labeled Lane Markers Using Maps, *ICCVW*, pp. 832-839.
- Benamara, N. K., Val-Calvo, M., Álvarez-Sánchez, J. R., Díaz-Morcillo, A., Ferrández-Vicente, J. M., Fernández-Jover, E. & Stambouli, T. B. (2020), Real-Time Facial Expression Recognition Using Smoothed Deep Neural Network Ensemble, *Integrated Computer-Aided Engineering*, **28**(1), 97-111.
- Chen, P., Wu, Y., Qin, Y., Yang, H. & Huang, Y. (2019), Rail Fastener Defect Inspection Based on Uav Images: A Comparative Study, *Proc. EITRT*, Qingdao, China, pp. 685-694.
- Chen, Z., Liu, Q. & Lian, C. (2019), Pointlanenet: Efficient End-to-End Cnns for Accurate Real-Time Lane Detection, *Proc. IV*, Paris, France, pp. 2563-2568.
- Cheng, C. S., Behzadan, A. H. & Noshadravan, A. (2021), Deep Learning for Post - Hurricane Aerial Damage Assessment of Buildings, *Computer-Aided Civil and Infrastructure Engineering*, **36**(6), 695-710.
- Cheng, Y., Zeng, J., Wang, Z. & Song, D. (2023), A Health State-Related Ensemble Deep Learning Method for Aircraft Engine Remaining Useful Life Prediction, *Applied Soft Computing*, **135**.
- Foudeh, H. A., Luk, P. C.-K. & Whidborne, J. F. (2021), An Advanced Unmanned Aerial Vehicle (Uav) Approach Via Learning-Based Control for Overhead Power Line Monitoring: A Comprehensive Review, *IEEE Access*, **9**, 130410-130433.
- Geng, Y., Pan, F., Jia, L., Wang, Z., Qin, Y., Tong, L. & Li, S. (2022), Uav-Lidar-Based Measuring Framework for Height and Stagger of High-Speed Railway Contact Wire, *IEEE Transactions on Intelligent Transportation Systems*, **23**(7), 7587-7600.
- Geng, Y., Wang, Z., Jia, L., Qin, Y., Chai, Y., Liu, K. & Tong, L. (2023), 3dgraphseg: A Unified Graph Representation-Based Point Cloud Segmentation Framework for Full-Range Highspeed Railway Environments, *IEEE Transactions on Industrial Informatics*, 1-13.
- Gibert, X., Patel, V. M. & Chellappa, R. (2017), Deep Multitask Learning for Railway Track Inspection, *IEEE Transactions on Intelligent Transportation Systems*, **18**(1), 153-164.
- Guclu, E., Aydin, I. & Akin, E. (2021), Development of Vision-Based Autonomous Uav for Railway Tracking, *Proc. 3ICT*, Institute of Electrical and Electronics Engineers Inc., Virtual, Online, Bahrain, pp. 120-125.
- Guo, F., Qian, Y., Wu, Y., Leng, Z. & Yu, H. (2021), Automatic Railroad Track Components Inspection Using Real - Time Instance Segmentation, *Computer-Aided Civil and Infrastructure Engineering*, **36**(3), 362-377.
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017), Mask R-Cnn, *ICCV*, pp. 2980-2988.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016), Deep Residual Learning for Image Recognition, *Proc. CVPR*, pp. 770-778.
- Hou, Y., Ma, Z., Liu, C. & Loy, C. C. (2021), Learning Lightweight Lane Detection Cnns by Self Attention Distillation, *Proc. ICCV*, Seoul, Korea, Republic of, pp. 1013-1021.
- Jiang, S. & Zhang, J. (2020), Real - Time Crack Assessment Using Deep Neural Networks with Wall - Climbing Unmanned Aerial System, *Computer-Aided Civil and Infrastructure Engineering*, **35**(6), 549-564.
- Kang, D. & Cha, Y.-J. (2018), Autonomous Uavs for Structural Health Monitoring Using Deep Learning and an Ultrasonic Beacon System with Geo-Tagging, *Computer-Aided Civil and Infrastructure Engineering*, **33**(10), 885-902.
- Kim, S., Kim, D., Jeong, S., Ham, J.-W., Lee, J.-K. & Oh, K.-Y. (2020), Fault Diagnosis of Power Transmission Lines Using a Uav-Mounted Smart Inspection System, *IEEE Access*, **8**, 149999-150009.
- Ko, Y., Lee, Y., Azam, S., Munir, F., Jeon, M. & Pedrycz, W. (2022), Key Points Estimation and Point Instance Segmentation Approach for Lane Detection, *IEEE Transactions on Intelligent Transportation Systems*, **23**(7), 8949-8958.
- Li, C., He, Q. & Wang, P. (2022), Estimation of Railway Track Longitudinal Irregularity Using Vehicle Response with Information Compression and Bayesian Deep Learning, *Computer-Aided Civil and Infrastructure Engineering*, **37**(10), 1260-1276.

- Li, X., Li, J., Hu, X. & Yang, J. (2020), Line-Cnn: End-to-End Traffic Line Detection with Line Proposal Unit, *IEEE Transactions on Intelligent Transportation Systems*, **21**(1), 248-258.
- Li, Y., Dong, H., Li, H., Zhang, X., Zhang, B. & Xiao, Z. (2020), Multi-Block Ssd Based on Small Object Detection for Uav Railway Scene Surveillance, *Chinese Journal of Aeronautics*, **33**(6), 1747-1755.
- Lin, T. Y., Goyal, P., Girshick, R., He, K. & Dollar, P. (2020), Focal Loss for Dense Object Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**(2), 318-327.
- Liu, J., Wang, Z., Wu, Y., Qin, Y., Cao, X. & Huang, Y. (2020), An Improved Faster R-Cnn for Uav-Based Catenary Support Device Inspection, *International Journal of Software Engineering and Knowledge Engineering*, **30**(07), 941-959.
- Liu, R., Yuan, Z., Liu, T. & Xiong, Z. (2021), End-to-End Lane Shape Prediction with Transformers, *Proc. WACV*, Virtual, Online, United States, pp. 3693-3701.
- Liu, Y. F., Nie, X., Fan, J. S. & Liu, X. G. (2020), Image - Based Crack Assessment of Bridge Piers Using Unmanned Aerial Vehicles and Three - Dimensional Scene Reconstruction, *Computer-Aided Civil and Infrastructure Engineering*, **35**(5), 511-529.
- Macias-Garcia, E., Galeana-Perez, D., Medrano-Hermosillo, J. & Bayro-Corrochano, E. (2021), Multi-Stage Deep Learning Perception System for Mobile Robots, *Integrated Computer-Aided Engineering*, **28**(2), 191-205.
- Martins, G. B., Papa, J. P. & Adeli, H. (2020), Deep Learning Techniques for Recommender Systems Based on Collaborative Filtering, *Expert Systems*, **37**(6), 1-21.
- McCoy, J., Rawal, A., Rawat, D. B. & Sadler, B. M. (2022), Ensemble Deep Learning for Sustainable Multimodal Uav Classification, *IEEE Transactions on Intelligent Transportation Systems*, 1-10.
- Narazaki, Y., Hoskere, V., Hoang, T. A., Fujino, Y., Sakurai, A. & Spencer, B. F. (2020), Vision - Based Automated Bridge Component Recognition with High - Level Scene Consistency, *Computer-Aided Civil and Infrastructure Engineering*, **35**(5), 465-482.
- Neven, D., De Brabandere, B., Georgoulis, S., Proesmans, M. & Van Gool, L. (2018), Towards End-to-End Lane Detection: An Instance Segmentation Approach, *Proc. IV*, Changshu, Suzhou, China, pp. 286-291.
- Oudshoorn, M., Koppenberg, T. & Yorke - Smith, N. (2022), Optimization of Annual Planned Rail Maintenance, *Computer-Aided Civil and Infrastructure Engineering*, **37**(6), 669-687.
- Pan, X., Shi, J., Luo, P., Wang, X. & Tang, X. (2018), Spatial as Deep: Spatial Cnn for Traffic Scene Understanding, *Proc. AAAI*, New Orleans, LA, USA, pp. 7276-7283.
- Philon, J. (2019), Fastdraw: Addressing the Long Tail of Lane Detection by Adapting a Sequential Prediction Network, *Proc. CVPR*, pp. 11574-11583.
- Qin, Z., Wang, H. & Li, X. (2020), Ultra Fast Structure-Aware Deep Lane Detection, *Proc. ECCV*, Cham, pp. 276-291.
- Rafiei, M. H. & Adeli, H. (2018), Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes, *Journal of Construction Engineering and Management*, **144**(12).
- Rafiei, M. H. & Adeli, H. (2018), A Novel Unsupervised Deep Learning Model for Global and Local Health Condition Assessment of Structures, *Engineering Structures*, **156**, 598-607.
- Ren, S., He, K., Girshick, R. & Sun, J. (2017), Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(6), 1137-1149.
- Sahebddivani, S., Arefi, H. & Maboudi, M. (2020), Rail Track Detection and Projection-Based 3d Modeling from Uav Point Cloud, *Sensors (Basel)*, **20**(18), 1-15.
- Saini, A. & Singh, D. (2021), Dronertef: Development of a Novel Adaptive Framework for Railroad Track Extraction in Drone Images, *Pattern Analysis and Applications*, **24**(4), 1549-1568.
- Tabelini, L., Berriel, R., Paixao, T. M., Badue, C., de Souza, A. F. & Oliveira-Santos, T. (2020), Polylandenet: Lane Estimation Via Deep Polynomial Regression, *Proc. ICPR*, Virtual, Milan, Italy, pp. 6150-6156.
- Tabelini, L., Berriel, R., Paixão, T. M., Badue, C., Souza, A. F. D. & Oliveira-Santos, T. (2021), Keep Your Eyes on the Lane: Real-Time Attention-Guided Lane Detection, *Proc. CVPR*, pp. 294-302.
- Tian, Y., Zhang, C., Jiang, S., Zhang, J. & Duan, W. (2021), Noncontact Cable Force Estimation with Unmanned Aerial Vehicle and Computer Vision, *Computer-Aided Civil and Infrastructure Engineering*, **36**(1), 73-88.
- Tong, L., Wang, Z., Jia, L., Qin, Y., Wei, Y., Yang, H. & Geng, Y. (2022), Fully Decoupled Residual Convnet for Real-Time Railway Scene Parsing of Uav Aerial Images, *IEEE Transactions on Intelligent Transportation Systems*, **23**(9), 14806 - 14819.
- TuSimple. (2017), Tusimple: Lane Detection Benchmark, <https://Github.Com/Tusimple/Tusimple-Benchmark>.
- V. Shajihan, S. A., Hoang, T., Mechitov, K. & Spencer, B. F. (2022), Wireless Smartvision System for Synchronized Displacement Monitoring of Railroad Bridges, *Computer-Aided Civil and Infrastructure Engineering*, **37**(9), 1070-1088.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017), Attention Is All You Need, *Proc. NIPS*, Long Beach, California, USA, pp. 6000-6010.
- Wang, Z., Wang, N., Zhang, H., Jia, L., Qin, Y., Zuo, Y., Zhang, Y. & Dong, H. (2023), Segmentalized Mmmr Features and Cost-Sensitive Elm with Fixed Inputs for Fault Diagnosis of High-Speed Railway Turnouts, *IEEE Transactions on Intelligent Transportation Systems*, 1-13.
- Weng, Y., Shan, J., Lu, Z., Lu, X. & Spencer, B. F. (2021), Homography - Based Structural Displacement Measurement for Large Structures Using Unmanned Aerial Vehicles, *Computer-Aided Civil and Infrastructure Engineering*, **36**(9), 1114-1128.
- Wu, Y., Qin, Y., Qian, Y., Guo, F., Wang, Z. & Jia, L. (2022), Hybrid Deep Learning Architecture for Rail Surface Segmentation and Surface Defect Detection, *Computer-Aided Civil and Infrastructure Engineering*, **37**(2), 227-244.
- Wu, Y., Qin, Y., Wang, Z. & Jia, L. (2018), A Uav-Based Visual Inspection Method for Rail Surface Defects, *Applied Sciences*, **8**(7), 1028-1047.
- Wu, Y., Qin, Y., Wang, Z., Ma, X. & Cao, Z. (2020), Densely Pyramidal Residual Network for Uav-Based Railway Images Dehazing, *Neurocomputing*, **371**, 124-136.
- Xu, B. & Liu, C. (2022), A 3d Reconstruction Method for Buildings Based on Monocular Vision, *Computer-Aided Civil and Infrastructure Engineering*, **37**(3), 354-369.
- Xu, H., Wang, S., Cai, X., Zhang, W., Liang, X. & Li, Z. (2020), Curvelane-Nas: Unifying Lane-Sensitive Architecture Search and Adaptive Point Blending, *Proc. ECCV*, Cham, pp. 689-704.
- Xu, J., Gui, C. & Han, Q. (2020), Recognition of Rust Grade and Rust Ratio of Steel Structures Based on Ensembled Convolutional Neural Network, *Computer-Aided Civil and Infrastructure Engineering*, **35**(10), 1160-1174.
- Yeum, C. M. & Dyke, S. J. (2015), Vision-Based Automated Crack Detection for Bridge Inspection, *Computer-Aided Civil and Infrastructure Engineering*, **30**(10), 759-770.
- Ying, L., Hoang, T., Haas, N., Otto, C. & Pankanti, S. (2014), Rail Component Detection, Optimization, and Assessment for Automatic Rail Track Inspection, *IEEE Transactions on Intelligent Transportation Systems*, **15**(2), 760-770.
- Yoo, S., Seok Lee, H., Myeong, H., Yun, S., Park, H., Cho, J. & Hoon Kim, D. (2020), End-to-End Lane Marker Detection Via Row-Wise Classification, *CVPRW*, Virtual, Online, USA, pp. 4335-4343.
- Zheng, T., Fang, H., Zhang, Y., Tang, W., Yang, Z., Liu, H. & Cai, D. (2021), Resa: Recurrent Feature-Shift Aggregator for Lane Detection, *Proc. AAAI*, pp. 3547-3554.
- Zheng, T., Huang, Y., Liu, Y., Tang, W. & Yang, Z. (2022), Clrnet: Cross Layer Refinement Network for Lane Detection, *Proc. CVPR*, pp. 888-897.
- Zhong, J., Liu, Z., Yang, C., Wang, H., Gao, S. & Nunez, A. (2022), Adversarial Reconstruction Based on Tighter Oriented Localization for Catenary Insulator Defect Detection in High-Speed Railways, *IEEE Transactions on Intelligent Transportation Systems*, **23**(2), 1109-1120.
- Zhuge, S., Xu, X., Zhong, L., Gan, S., Lin, B., Yang, X. & Zhang, X. (2022), Noncontact Deflection Measurement for Bridge through a Multi - Uavs System, *Computer-Aided Civil and Infrastructure Engineering*, **37**(6), 746-761.