drones

MDPI

Article

Railway Tracks Extraction from High Resolution Unmanned Aerial Vehicle Images Using Improved NL-LinkNet Network

Jing Wang [1,2], Xiwei Fan [1,2,*], Yunlong Zhang [3,*], Xuefei Zhang [4], Zhijie Zhang [5], Wenyu Nie [1,2], Yuanmeng Qi [1,2] and Nan Zhang [1,2]

[1] Key Laboratory of Seismic and Volcanic Hazards, China Earthquake Administration, Beijing 100029, China; jingwang@ies.ac.cn (J.W.); niewenyu0528@163.com (W.N.); qiyuanmeng523079@gmail.com (Y.Q.); zhangnan@ies.ac.cn (N.Z.)
[2] Institute of Geology, China Earthquake Administration, Beijing 100029, China
[3] China Railway Design Corporation, Tianjin 300308, China
[4] Land Satellite Remote Sensing Application Center, Ministry of Natural Resources, Beijing 100034, China; zhangxf@lasac.cn
[5] Institute of Strategic Planning, Chinese Academy of Environmental Planning, Beijing 100012, China; zhangzj@caep.org.cn
* Correspondence: fanxiwei@ies.ac.cn (X.F.); 13115311@bjtu.edu.cn (Y.Z.)

Abstract: The accurate detection of railway tracks from unmanned aerial vehicle (UAV) images is essential for intelligent railway inspection and the development of electronic railway maps. Traditional computer vision algorithms struggle with the complexities of high-precision track extraction due to challenges such as diverse track shapes, varying angles, and complex background information in UAV images. While deep learning neural networks have shown promise in this domain, they still face limitations in precisely extracting track line edges. To address these challenges, this paper introduces an improved NL-LinkNet network, named NL-LinkNet-SSR, designed specifically for railway track detection. The proposed NL-LinkNet-SSR integrates a Sobel edge detection module and a SimAM attention module to enhance the model's accuracy and robustness. The Sobel edge detection module effectively captures the edge information of track lines, improving the segmentation and extraction of target edges. Meanwhile, the parameter-free SimAM attention module adaptively emphasizes significant features while suppressing irrelevant information, broadening the model's perceptual field and improving its responsiveness to target areas. Experimental results show that the NL-LinkNet-SSR significantly outperforms the original NL-LinkNet model across multiple key metrics, including a more than 0.022 increase in accuracy, over a 4% improvement in F1-score, and a more than 3.5% rise in mean Intersection over Union (mIoU). These enhancements suggest that the improved NL-LinkNet-SSR offers a more reliable solution for railway track detection, advancing the field of intelligent railway inspection.

Keywords: deep learning; edge detection; railway track detection; attention mechanism

Citation: Wang, J.; Fan, X.; Zhang, Y.; Zhang, X.; Zhang, Z.; Nie, W.; Qi, Y.; Zhang, N. Railway Tracks Extraction from High Resolution Unmanned Aerial Vehicle Images Using Improved NL-LinkNet Network. *Drones* **2024**, *8*, 611. https://doi.org/10.3390/drones8110611

Academic Editor: Diego González-Aguilera

Received: 29 August 2024
Revised: 21 October 2024
Accepted: 23 October 2024
Published: 25 October 2024

1. Introduction

Railway transportation plays a vital role in national economic growth, serving as a key component of the transportation network. Precise railway track extraction is fundamental for creating detailed electronic maps, maintaining smooth operations, and safeguarding lives and property [1]. Traditionally, railway tracks have been inspected manually or with specialized vehicles, but these methods are often inefficient and lack comprehensive, high-frequency monitoring capabilities [2,3]. Manual inspections across extensive railway networks are highly inefficient and labor-intensive. Moreover, optical sensors on inspection vehicles often collect unsatisfactory data due to the loss of information details, adversely affecting the accuracy of subsequent defect analysis and detection. With advancements in

remote sensing and UAV technology, UAV inspections are increasingly utilized for the intelligent detection of railway tracks, significantly enhancing the safety of high-speed railways. Equipped with high-definition cameras, UAVs are able to efficiently collect data on railway infrastructure without disrupting normal operations, thus improving both detection accuracy and efficiency [4]. However, challenges remain due to the complex background information in UAV images, the diversity of railway track shapes, and the variability of shooting angles, which make high-precision extraction of track lines particularly difficult.

Traditional methods for railway track extraction often necessitate substantial prior knowledge from researchers. Several studies have utilized models for track detection using conventional computer vision algorithms [5,6]. However, these methods frequently misclassify railway tracks due to their spectral similarities to other features such as buildings, fields, water bodies, and parking lots. As a result, they often suffer from lower-than-expected classification accuracy and issues related to extraction precision and robustness. Therefore, it is essential that advanced detection algorithms be developed to improve the effectiveness of UAV-based railway track inspections.

Traditional machine learning techniques play a key role in image feature detection; techniques such as Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) provide robust solutions for detecting and describing local features in images under varying conditions [7,8]. Histogram of Oriented Gradients (HOG) and Haar cascades are also widely used in object detection as they capture shape information and rapidly recognize features [9,10]. These methods have established the groundwork for detecting features of railway tracks in UAV imagery. Deep learning has undergone rapid advancements in recent years, significantly improving image recognition and object detection from UAV imagery [11]. It is worth noting that the UAV benchmark study on object detection and tracking has been instrumental in driving these advancements, establishing foundational benchmarks and performance metrics for various applications of UAV-based visual systems [12,13]. This progress extends to various applications, including road extraction [14], lane detection [15], railway foreign body detection [16], rail surface defect inspection [17], and crop row detection and guidance system [18,19]. Deep learning-based methods for detecting railway tracks from high-resolution UAV imagery offer notable advantages over traditional approaches. They require less prior knowledge, reduce researchers' workload, and more effectively handle the complex environments encountered in railway track extraction. These methods have already been employed in extracting railway tracks from UAV imagery and in creating electronic railway maps [20,21].

Given the similarities between road extraction and railway track detection, methods developed for road extraction are also effective for railway track detection using UAV imagery. These methods identify the network structure of roads or railways in high-resolution remote sensing images (with a resolution of 0.5 to 1 m), sourced from unmanned aerial vehicle (UAV) remote sensing and satellite remote sensing platforms, by recognizing multi-scale features. Unlike unsupervised learning, which often relies on color-based segmentation, deep learning techniques utilize a range of features, including texture, geometric shapes, and line patterns, to extract roads [22]. Currently, many road detection algorithms for high-resolution remote sensing data utilize deep neural network models based on encoder-decoder structures such as FCN, UNet, and DeepLabV3+. Researchers are continually optimizing network architectures, objective functions, and training strategies to achieve more precise road segmentation.

Research on road extraction focused on enhancing backbone networks, context information extraction, and attention mechanisms. Convolutional neural networks (CNNs), as a foundational deep learning architecture, significantly contribute to road segmentation. For example, Tao et al. proposed the Seg-Road model, which combines Transformer and CNN techniques for road extraction from remote sensing images [23]. Similarly, Qiu L developed the Semantic Geometry Network (SGNet), which uses dual-branch backbones to extract roads from high-resolution images [24]. Unlike CNN models that use dense layers to generate fixed-length feature vectors and require fixed-size images, Fully Convolutional

Networks (FCNs) employ interpolation layers to upsample feature maps, allowing them to restore the input size and process images of any dimension. Varia et al. utilized the FCN-32 network to segment road sections from ultra-high-resolution UAV imagery [25], while Kestur et al. developed a novel U-shaped FCN (UFCN) specifically designed for road extraction from UAV images [26]. Zhu et al. introduced the Global Context-Aware and Batch-Independent Network (GCB-Net) for this purpose [27]. In GCB-Net, Global Context-Aware (GCA) blocks within the encoder enhance the capture of global spatial relationships, while multi-parallel unfolding convolutions enable the extraction of multi-scale road features, thereby improving the model's overall efficacy and connectivity of road topology. Additionally, Dai L et al. introduced the Road-Enhanced Deformable Attention Network (RADANet), which leverages road shape priors and deformable attention mechanisms to extract road information from high-resolution images. This method effectively captures semantic shape information and long-range dependencies between road features [28].

Despite the substantial advancements in road segmentation through deep learning, the application of these methods to railway track extraction poses several challenges. Existing models often necessitate enhanced edge detection accuracy for track lines and face difficulties with the linear structural characteristics of railway tracks. Deep learning-based pixel-level feature extraction may introduce noise, gaps, or discontinuities, and these models fail to adapt to variations in the tilt angles of railway tracks in UAV imagery, which reduces segmentation accuracy. Additionally, deep neural networks are often computationally inefficient due to their multiple layers and extensive parameter sets. Another challenge is the absence of specialized training datasets specifically designed for railway track extraction.

To address these challenges, this paper introduces an improved NL-LinkNet deep learning network, termed NL-LinkNet-SSR, specifically designed for extracting railway tracks from UAV aerial images [29]. This network provides a robust solution for the automated extraction of railway track lines. Building on the conventional NL-LinkNet architecture, the model integrates a Sobel edge detection module and a parameter-free SimAM attention mechanism. These enhancements markedly enhance the network's ability to detect railway track edges, thus increasing the precision and reliability of the track extraction process. The key contributions of this study include:

(1) The encoder integrates Sobel edge detection modules and non-local blocks to effectively extract edge information of railway tracks from the input images and incorporate it with the original feature maps. This integration improves the network's edge perception capabilities, enabling the model to capture fine details and contextual information about the railway tracks, thus improving extraction accuracy and robustness.

(2) The decoder incorporates the SimAM attention mechanism, which is applied to the output feature maps of each decoder block. This results in weighted feature maps that emphasize the railway track regions, selectively amplifying the feature responses in these areas. The parameter-free nature of SimAM ensures high computational efficiency without the need for additional learning parameters.

(3) A new dataset consisting of 12,130 high-resolution railway track images and their corresponding label images has been developed, providing a valuable data resource for railway track extraction from UAV images.

The remainder of this paper is organized as follows: Section 2 reviews related work on deep learning-based railway track extraction methods. Section 3 introduces the experimental data from UAV images, the algorithm framework for track extraction, and the network structure. Section 4 describes the experimental setup and evaluation metrics. Section 5 presents a detailed analysis of the experimental results, including comparisons of different models and ablation studies. Finally, Sections 6 and 7 provide the discussion and conclusions, respectively.

2. Related Work

Tong et al. [20] introduced a novel anchor-adaptive dual-branch architecture (DBA) called ARTNet, based on the Progressive Learning Detector (PLD), for railway track detection from UAV aerial images, which significantly enhances the robustness of railway track extraction. Several railway track extraction frameworks utilize an encoder-decoder structure to integrate multi-layer features of convolutional neural networks (CNNs), effectively leveraging multi-scale information across different semantic levels. For example, Mammeri et al. [30] employed a U-Net network with an encoder-decoder structure to extract railway areas from drone images. Similarly, Weng et al. [31] proposed an improved method for railway track extraction using the DeepLabV3+ model, which effectively eliminated errors such as holes and spots in the extracted track lines through the use of morphological algorithms. Additionally, Weng et al. [21] developed an enhanced D-LinkNet convolutional neural network that integrates a specifically designed edge detection module to fuse multi-level features, thereby improving the model's ability to segment and extract track edges. Tu et al. [32] proposed the RT-GAN framework, based on a generative adversarial network structure, for precise railway track segmentation in drone images. Despite these advancements in railway track area extraction and railway object recognition, there has been no research specifically targeting the extraction of railway track lines from UAV images. Furthermore, current methods still exhibit limitations in accuracy and robustness when extracting railway track lines, especially in scenes with varying tilt angles and complex background noise. This underscores the need for a railway track detection method that is better suited to such challenging conditions.

3. Dataset and Methodology

The algorithmic framework outlined in this paper is depicted in Figure 1, which significantly enhances the detection of railway tracks from aerial images captured by drones. The core innovation lies in the advanced adaptation of the NL-LinkNet network, which now incorporates a Sobel edge detection module and a SimAM attention mechanism residual module [31]. These enhancements are specifically designed to improve the network's sensitivity to railway track features, which is a critical aspect for achieving higher precision in identifying and delineating these structures from high-resolution drone imagery. Initially, high-resolution images of ground railway tracks were captured using drone aerial photography, and corresponding labeled images were generated. Data augmentation techniques were applied to enhance model training. The NL-LinkNet network was subsequently enhanced through the integration of an edge detection module and a SimAM attention mechanism residual module [33], aimed at improving the model's ability to identify railway track features. The upgraded NL-LinkNet network was then utilized for training and testing to further assess the accuracy of the model's predictions. Finally, comparative experiments and ablation studies were conducted to validate the effectiveness and advantages of the model.

3.1. UAV Data Preprocessing and Datasets Introduction

This dataset was acquired through aerial photography with DJI (Da-Jiang Innovation) drones, specifically using the DJI Matrice 300 RTK model equipped with a Zenmuse P1 camera, covering specific railway lines in Qingdao and Nanjing, China. The drone operates at a typical altitude of 150 m with a pixel size of 4.4 μm and a 35 mm lens, resulting in a Ground Sample Distance (GSD) of 15.08 cm, which ensures detailed and precise imagery for analysis. This task includes 214 high-resolution images, each with a resolution of 8192 × 5460 pixels at 96 dpi, totaling approximately 49.33 GB. The images were captured along a predetermined linear flight path at regular intervals, which introduced some redundancy in the railway information.

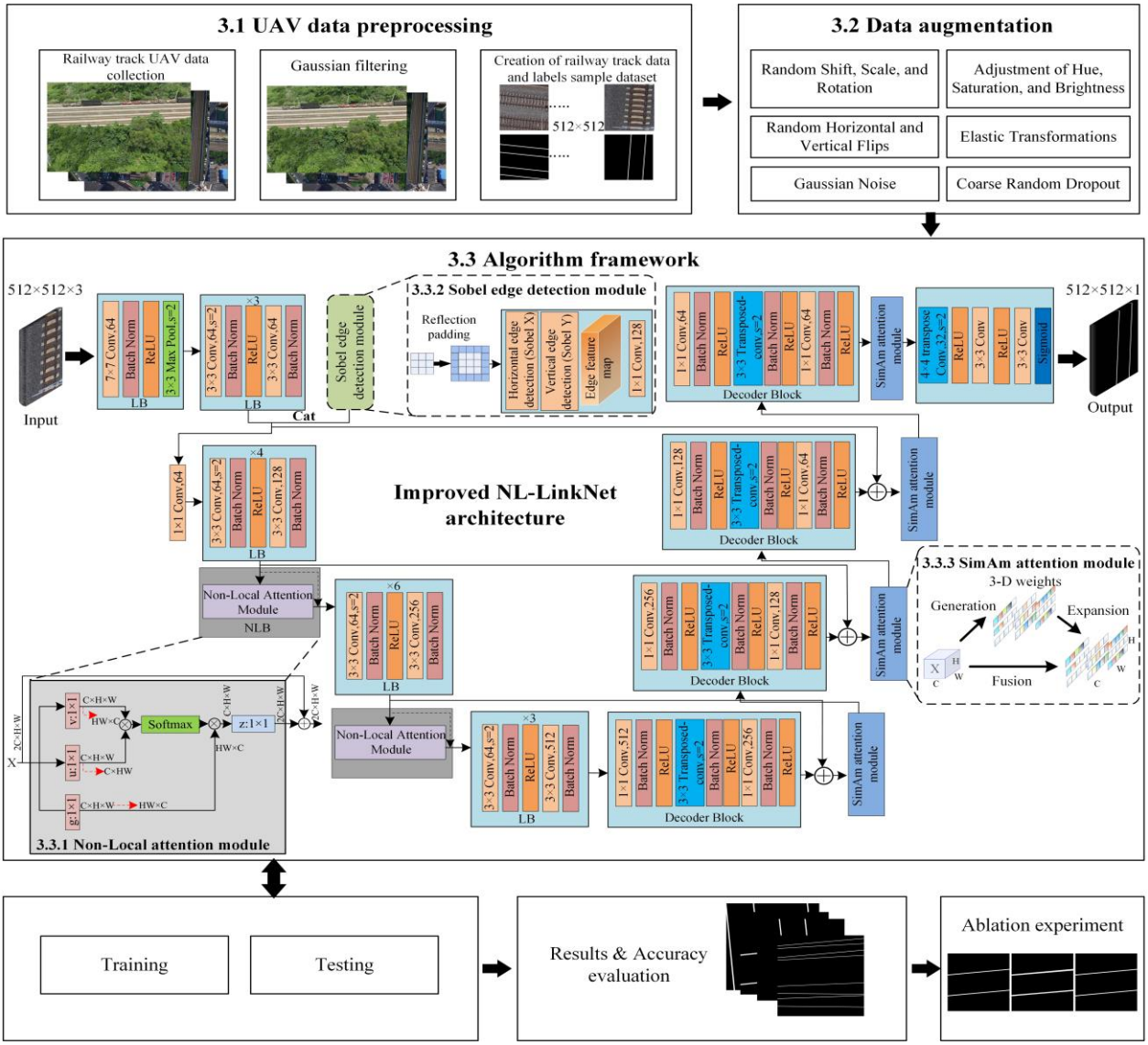

Figure 1. Methodological framework of the overall methods used in this study. This figure depicts content related to Sections 3.1–3.3 and Sections 3.3.1–3.3.3 in the text.

Images containing railway tracks were specifically selected to ensure complexity and diversity, meeting the requirements for railway track line extraction tasks. These images were initially preprocessed with Gaussian filtering to reduce noise. Using the coordinate information from the drone images, track line points were calculated through linear interpolation, and these points were connected to form lines, resulting in binarized label images with the track lines. These label images were automatically generated using a Python program.

For model training, the filtered images and corresponding binarized track line label images were cropped using a sliding window algorithm, resulting in 12,130 image slices. These slices were then split into a training set and a test set in a 9:1 ratio, where the test set also served as a validation set. The original images and label images of part of the dataset can be seen in Figure 2.

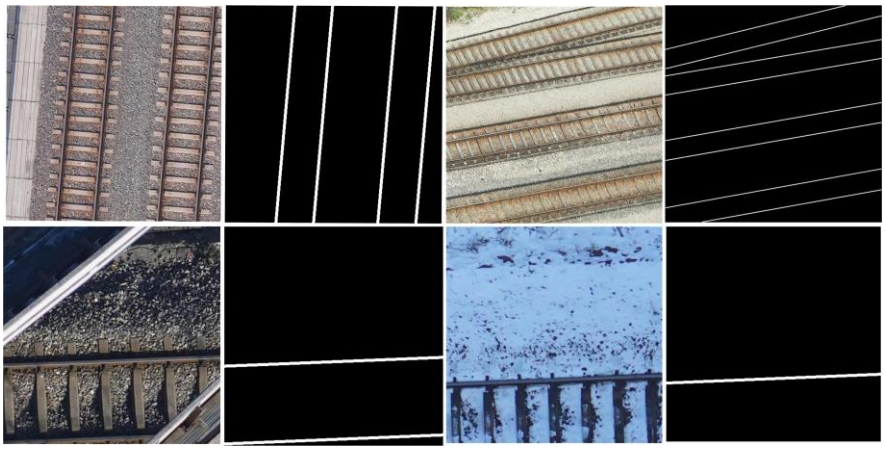

Figure 2. Original and labeled images of railway track lines.

3.2. Data Augmentation

Several data augmentation techniques were applied to the railway track image datasets to enhance the model's generalization and robustness. These methods included random adjustments of image hue, saturation, and brightness to simulate various lighting conditions, with hue altered within a -60 to 60-degree range and saturation and brightness varied between -50 and 50 units. Perspective transformations introduced variability through random translations, scaling, and rotations, with shift limits from -0.7 to 0.7, scale limits from -0.8 to 0.8, and rotation limits from -90 to 90 degrees [34]. Gaussian noise was added with a variance limit ranging from 15 to 60 to simulate sensor noise and realistic image acquisition conditions. Elastic transformations with parameters such as alpha of 120, a sigma of 6, and an alpha affine of 3.6 simulated natural scene distortions to aid the model in adapting to local deformations. Coarse Dropout randomly created black patches or holes, ranging from 2 to 8 holes of 8×8 to 16×16 pixels, in order to simulate occlusions. Additionally, random horizontal and vertical flips and 90-degree rotations were employed to further increase the diversity and robustness of the dataset against directional biases. All these augmentations were implemented using the Albumentations library, ensuring a robust and diverse set of images that trains the model to perform effectively under varied conditions.

3.3. Algorithm Framework

The improved NL-LinkNet network architecture is illustrated in Figure 1. This framework is based on a ResNet34 [35] encoder-decoder structure, designed to enhance feature representation in railway track line segmentation tasks. The network incorporates non-local attention modules in the third and fourth encoder layers to capture long-range dependencies and global contextual information.

A Sobel edge detection module is also integrated into the network to extract edge features of railway track lines from the images. The module is a key image processing technique that calculates the gradient magnitude at each pixel to highlight areas where intensity sharply changes, indicating edges or boundaries within the image. The Sobel edge detection module uses horizontal and vertical Sobel filters for edge detection in both orientations, applying them via convolutional operations to emphasize relevant intensity changes. The combined edge maps from both directions create a comprehensive scene depiction, further refined by a 1×1 convolution that adjusts channel dimensions to match subsequent network layers. This integration within the deep learning framework significantly enhances the model's ability to detect railway tracks with high accuracy. These edge features are then fused with the output from the first encoder layer through a channel adjustment layer, thereby enhancing the model's capability to extract detailed track line features.

Additionally, a SimAM attention module is appended to each decoder block to refine feature representation through an adaptive mechanism. The module is an attention mechanism that dynamically adjusts the focus of the neural network on important features within an image. By computing attention scores based on the significance of each feature, SimAM effectively directs the model's computational resources towards areas of interest, enhancing detection accuracy and efficiency. This method is particularly beneficial in environments with variable and intricate backgrounds where distinguishing key features from noise is critical.

After several convolutional layers and nonlinear activation functions, the network produces high-precision segmentation results for railway track lines. By leveraging the strong feature extraction capabilities of the ResNet34 network and incorporating these enhancement modules, the architecture achieves superior performance in extracting railway track lines from complex scenes.

3.3.1. Non-Local Attention Module

Nonlocal attention blocks (NLBs) are a key enhancement for convolutional neural networks (CNNs), designed to capture long-range dependencies in feature maps [29,36]. Traditional CNNs often face limitations due to their restricted receptive fields, which hinder their ability to reference distant spatial information. Nonlocal blocks overcome this limitation by calculating the response at a given position as a weighted sum of features from all positions in the input feature map. This approach allows each spatial point to aggregate contextual information from the entire image, enhancing the network's ability to process and understand complex patterns that span large areas.

In railway track extraction, nonlocal operations provide significant benefits. High-resolution satellite images of railway tracks may be obscured by shadows, trees, or buildings, making accurate detection challenging for conventional methods. By incorporating the NLBs, models can leverage global information across the entire image, improving the precision of track extraction even in the presence of such obstacles. The ability of nonlocal blocks to compute a weighted sum across the entire feature map enables each spatial point to gather contextual information from the whole image. This capability enhances the model's performance, as demonstrated by NL-LinkNet's superior results in the DeepGlobe Challenge, where it outperformed state-of-the-art models with fewer parameters and faster training times.

The function of the nonlocal block can be described by the following equation:

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \tag{1}$$

where y_i is the output at position i, and x_i and x_j represent the input features at positions i and j, respectively. The function $f(x_i, x_j)$ computes the pairwise relationship (or affinity) between features at positions i and j. $C(x)$ is a normalization factor, typically set as $C(x) = \sum_{\forall j} f(x_i, x_j)$. A common choice for f is the embedded Gaussian function:

$$f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)} \tag{2}$$

where $\theta(x_i) = W_\theta x_i$ and $\phi(x_j) = W_\theta x_j$ are linear embeddings, and the function $g(x_j) = W_g x_j$ is a linear embedding applied to the input features. The non-local block uses learnable weights W_θ, W_ϕ, W_g to transform the linear embeddings of the input features. Then, it calculates the pairwise function of similarity between features using the embedded Gaussian function. It aggregates features from all positions weighted by the calculated similarity and adds the aggregated features back to the original input to form the output, maintaining residual connections that preserve both local and global information.

3.3.2. Edge Detection Module

The Sobel operator is a commonly used method for edge detection in image processing. It utilizes two 3×3 kernels: one for detecting horizontal edges and another for detecting vertical edges. These kernels convolve with the input image to approximate the gradients in the horizontal and vertical directions. The gradient magnitude at each pixel is then computed as the square root of the sum of the squares of these horizontal and vertical gradients. In the model, the feature maps generated by the Sobel edge detection module are concatenated with those from the first layer of the LB module in the encoder along the channel dimension. This process enriches the model's representation of the input image.

During implementation, the Sobel kernels are first transferred to the same GPU device as the input tensor to ensure consistent computation. The input tensor is then padded using reflection padding to handle boundary pixels and preserve edge information. Assuming the input tensor x has dimensions (C, H, W), where C is the number of channels, H is the height, and W is the width, the padded tensor dimensions become $(C, H+2, W+2)$. The Sobel operator, with its 3×3 kernel, is then applied in both the horizontal and vertical directions to the input tensor, producing two edge maps that represent the horizontal and vertical edge intensities. The convolved output tensor size remains (C, H, W) for both the horizontal and vertical convolutions, thereby maintaining the same number of channels as the input.

To compute the overall edge strength (gradient magnitude) for each pixel, the square root of the sum of the squares of the horizontal and vertical edge maps is calculated. A small epsilon (1e-8) is added to this calculation to avoid zero gradients. The resulting edge strength is then maximized along the channel dimension while maintaining the original spatial dimensions, resulting in a feature map of size $(1, H+2, W+2)$. Finally, a 1×1 convolution layer is applied to adjust the channel dimensions of the edge map, ensuring it matches the expected input size of the subsequent layers. This module effectively converts the input image into an edge map, highlighting edges in both horizontal and vertical directions, thereby providing crucial information for further railway track line processing tasks.

3.3.3. SimAM Attention Mechanism

SimAM is a lightweight, parameter-free attention mechanism designed for convolutional neural networks and is commonly used in visual tasks such as image classification, object detection, and image segmentation [33]. Unlike traditional channel or spatial attention modules, SimAM calculates three-dimensional attention weights for feature maps directly within the inference layer without increasing the network's parameter count. Inspired by neuroscience principles, this module employs an energy function to evaluate the importance of each neuron, thereby reducing model complexity and computational cost. This approach enhances the representational capacity of convolutional neural networks, leading to improved performance in visual tasks, such as railway track line extraction.

First, the module calculates the mean $\hat{\mu}$ and variance $\hat{\sigma}^2$ of the input feature map X. The energy function for each neuron is defined as

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \lambda\right)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{3}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the mean and variance of the feature map, respectively, and λ is a predefined coefficient. This equation indicates that a lower energy value e_t^* signifies a more important neuron for visual processing. Consequently, the importance of each neuron can be determined by $\frac{1}{e_t^*}$. The squared difference (d) between each feature and the mean is then calculated and normalized by the adjusted variance term (v), given by $v = \frac{d.sum(dim=[2,3])}{n}$, where n is the number of elements minus one. The inverse of the energy function E_{inv} is computed as $E_{inv} = \frac{d}{4(v+\lambda)} + 0.5$. Finally, the attention map is generated by applying

a sigmoid function to the inverse energy values and refining the feature map by scaling, $\widetilde{X} = sigmod\left(\frac{1}{E}\right) \cdot X$.

4. Experiments

4.1. Implementation Details

The experiments were conducted on a Windows 11 operating system, utilizing an NVIDIA RTX 4090 graphics card with 24 GB of VRAM and CUDA version 11.2. The PyTorch framework was used for model training. Several optimization algorithms, including SGD, RMSprop, Adam, and AdamW, were tested during the experiments. AdamW was ultimately chosen due to its superior performance in terms of faster convergence, improved training stability, and better generalization. SGD showed slower convergence on complex data, while RMSprop and Adam offered adaptive learning but occasionally led to instability or overshooting. AdamW, by decoupling weight decay from the optimization process, provided faster convergence and better generalization and was proven to be the most stable and effective optimizer in our experiments [37]. A batch size of 8 was chosen to balance computational efficiency and memory usage. The initial learning rate was set to 2e-5, and a weight decay of 1e-4 was applied to mitigate overfitting. A learning rate decay strategy was employed using the ReduceLROnPlateau scheduler, which reduced the learning rate by a factor of 0.8 after 10 consecutive epochs without improvement in validation loss. The model was trained for 100 epochs to ensure thorough learning and convergence.

4.2. Loss Function

The loss function used in this study for railway track line detection is a combination of Dice loss and Focal loss, referred to as Dice Focal Loss [38]. This hybrid loss function effectively addresses the significant class imbalance between positive and negative samples in the dataset.

Dice loss measures the overlap between predicted and true binary masks [39], with a focus on accurately predicting positive samples. It is computed using the Dice coefficient. The Dice loss is given by:

$$\mathrm{D}_f = 1 - \frac{2\sum y_{true} y_{pred} + \epsilon}{\sum y_{true} + \sum y_{pred} + \epsilon} \tag{4}$$

where y_{true} and y_{pred} are the ground truth and predicted binary masks, respectively, and ϵ is a small constant to prevent division by zero.

The Focal loss is utilized to further mitigate the issue of class imbalance by focusing more on hard-to-classify samples [40]. It modifies the standard binary cross-entropy loss by introducing a modulating factor that decreases the loss contribution from easy examples and increases it for hard examples. The Focal loss is defined as:

$$F = -\alpha(1 - p_t)^{\gamma} \log(p_t) \tag{5}$$

where p_t is the predicted probability of the true class, α is a balancing factor, and γ is a focusing parameter that adjusts the rate at which easy examples are down-weighted.

By combining these two loss functions, the Dice Focal Loss effectively harnesses the strengths of both, providing robust performance in handling class imbalance and improving the model's ability to accurately detect railway tracks. The overall loss function is expressed as:

$$D_F = \mathrm{D}_f + F \tag{6}$$

4.3. Evaluation Metrics

In this study, several evaluation metrics were employed to assess the performance of the railway track line detection model, including accuracy, mean Intersection over Union (mIoU), precision, F1-score, recall, and the kappa coefficient. These metrics provide a

comprehensive evaluation of the model's effectiveness. All these indicators are derived from the confusion matrix. The definitions of each parameter in the confusion matrix are presented in Table 1.

Table 1. Confusion matrix diagram.

		Prediction		
		Railway Track Line	**Non-Railway Track Line**	**Sum**
Ground truth	**Railway track line**	TP	FN	TP + FN
	Non-railway track line	FP	TN	FP + TN
	Sum	TP + FP	FN + TN	TP + TN + FP + FN

Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted samples to the total number of samples. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{7}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Mean Intersection over Union (MIoU) evaluates the average overlap between the predicted and ground truth segments across all classes. It is calculated as:

$$\text{MIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{A_i \cap B_i}{A_i \cup B_i} \tag{8}$$

where N is the number of classes, Ai is the predicted set for class i, and Bi is the ground truth set for class i.

Recall (or sensitivity) calculates the proportion of true positive predictions among all actual positive samples, reflecting the model's ability to detect positive samples. It is given by:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

F1-score is the harmonic mean of precision and recall, providing a single metric that balances both precision and recall. It is defined as:

$$F_{1-\text{score}} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{10}$$

Kappa coefficient (Cohen's kappa) measures the agreement between the predicted and ground truth labels, adjusted for the agreement occurring by chance. It is calculated as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{11}$$

where p_0 is the observed agreement ratio, and p_e is the expected agreement by chance.

$$p_0 = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$p_e = \frac{(TP + FP) \times (TP + FN) + (TN + FP) \times (TN + FN)}{(TP + TN + FP + FN)^2} \tag{13}$$

5. Results and Analysis

5.1. Visualization of Railway Track Lines Extraction

The visualization results of the railway track line extraction are presented in Figure 3. These results demonstrate the model's strong generalization capability on the railway tracks dataset. It is clear that the proposed model effectively identifies railway track line regions across various angles and complex backgrounds. The model robustly identifies and segments railway track lines, effectively differentiating them from diverse environments, including snowy conditions, urban clutter, and amidst foliage. This is evidenced by the distinct contrast between the original images and the processed results, where track lines are prominently highlighted. These examples highlight the model's strong generalization capability across a broad range of scenarios, affirming its adaptability to different railway track orientations and environmental conditions. The implementation of the Sobel edge detection module, combined with the SimAM attention mechanism, enhances the model's sensitivity to subtle edge details, crucial for accurate track delineation in complex scenes. Furthermore, the inclusion of NLBs in the model architecture allows it to capture and utilize global contextual information, thus enabling the model to maintain performance even when local visual information is compromised by occlusions or blending with background textures. This advanced integration of NLBs with edge detection and attention mechanisms proves particularly effective in scenarios where track lines are obscured or blend with backgrounds of similar textures. It demonstrates the model's superior feature extraction capabilities and robustness, ensuring reliable track detection even under challenging real-world conditions.

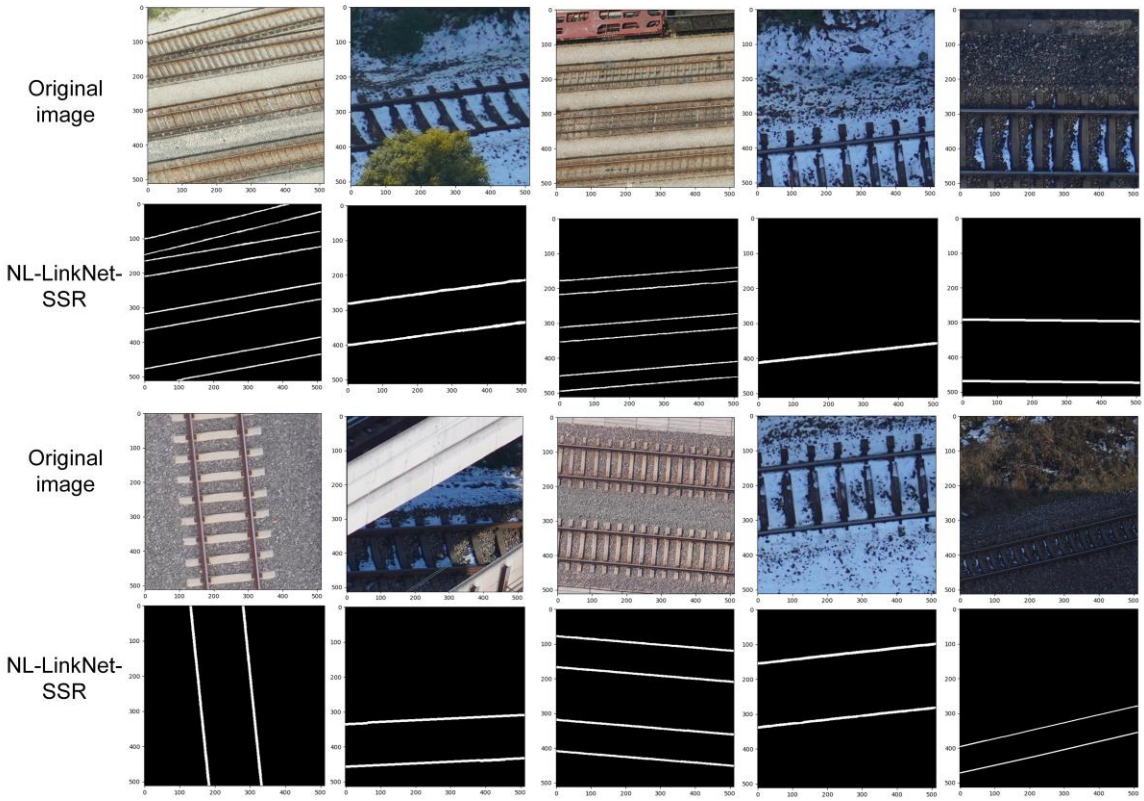

Figure 3. Railway track line extraction results based on the improved NL-LinkNet model (NL-LinkNet-SSR).

To demonstrate the effectiveness and robustness of the method in extracting railway track lines, the model's performance was compared against four other network models: NL-LinkNet, DeepLabv3+, U-Net, and FCN. All experiments were conducted under identical

conditions using the same dataset to ensure fairness and objectivity. The results of different models in railway track line extraction are illustrated in Figure 4.

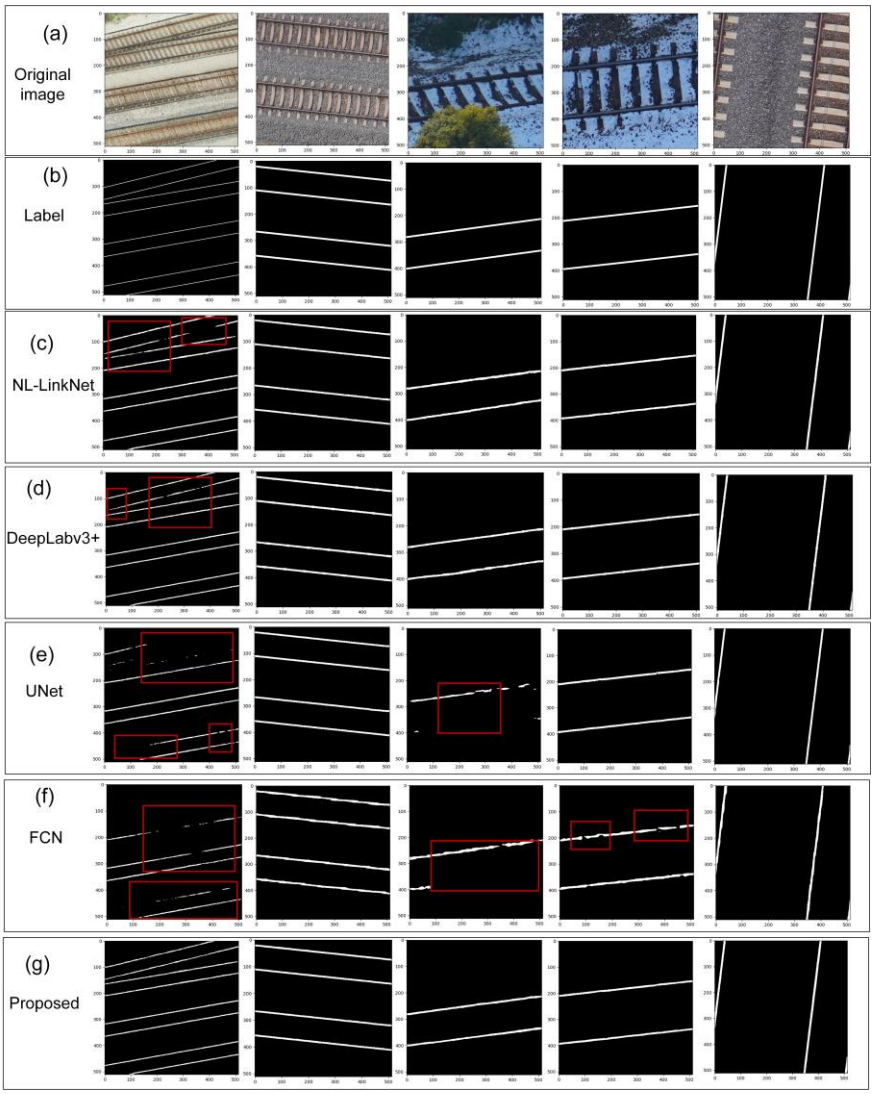

Figure 4. Comparative analysis of railway track line extraction results from different models. (**a**) UAV imagery; (**b**) ground truth; (**c**) NL-LinkNet; (**d**) Deeplabv3+; (**e**) U-Net; (**f**) FCN; and (**g**) The proposed method.

The NL-LinkNet model shows relatively accurate track line extraction but exhibits deviations and broken lines, especially in the first and third images, where the continuity of the extracted lines is compromised. The DeepLabv3+ model performs well in capturing most railway lines accurately, but there are minor inconsistencies, particularly in the first and third images, where the lines are not as precise as the ground truth. The U-Net model struggles with accurately extracting railway lines, displaying noticeable gaps and noise, especially in the first, second, and fourth images. The FCN model has difficulty maintaining the continuity of railway lines, resulting in significant gaps and misdetections, especially in the first, third, and fourth images. In contrast, the proposed method outperforms the other models, delivering the most accurate and continuous railway track line extractions. The extracted lines closely match the ground truth with minimal deviations and noise, demonstrating the robustness and effectiveness of the proposed approach. Moreover, compared to the proposed method, the U-Net, DeepLabv3+, and FCN models often miss segments with similar colors and textures and struggle to extract smaller segments. These models perform poorly in extracting railway track lines, particularly in shadowed areas,

due to insufficient feature learning when dealing with similar colors, textures, and shadows, leading to a loss of detail.

Overall, the proposed method demonstrates superior performance in accurately extracting railway track lines compared to NL-LinkNet, DeepLabv3+, U-Net, and FCN. This highlights the effectiveness of the proposed improvements in enhancing the precision and reliability of railway track detection from UAV imagery.

Table 2 presents the quantitative extraction metrics for railway track lines on the test set. The proposed model achieved the best performance with an accuracy (Acc) of 98.2%, an F1 score of 74.9%, a mean Intersection over Union (mIoU) of 65.3%, and a Kappa coefficient of 84.1%. The F1 score, which is the weighted average of precision and recall, serves as an objective measure of the model's performance. The model's F1 score of 74.9% outperforms the other models, with DeepLabv3+ following closely at 72.8%, which is 2.1% lower than the proposed model. In contrast, the FCN model performed the least effectively, with an F1 score of only 66.3%, which is 8.6% lower than the proposed model.

Table 2. The evaluation metrics of proposed and comparative models on railway track extraction tasks.

Method	Accuracy	F1-Score	mIoU	Recall	Kappa
NL-LinkNet	0.960	0.706	0.618	0.733	0.810
DeepLabv3+	0.974	0.728	0.636	0.762	0.842
UNet	0.959	0.694	0.603	0.745	0.740
FCN	0.964	0.663	0.595	0.727	0.768
Proposed	0.982	0.749	0.653	0.780	0.841

To validate the performance of the proposed method in railway track line extraction, the training loss convergence process of different models on the railway dataset was visualized. During training, the training loss typically decreases gradually, while the validation loss, computed on data outside the training set, assesses the model's generalization ability. A decrease in both losses indicates effective learning by the model.

The training and validation loss curves for various models, shown in Figure 5, provide insights into their performance and generalization capabilities over the training epochs. All models exhibit a steady decrease in training loss, reflecting effective learning. However, validation loss shows variability, highlighting the models' ability to generalize to unseen data. The NL-LinkNet (Figure 5a) and DeepLabv3+ (Figure 5b) models demonstrate a stable decrease in training loss, with validation loss following a similar trend, suggesting good generalization. However, the validation loss stabilizes at a higher value than the training loss, indicating some degree of overfitting. The UNet (Figure 5c) and FCN (Figure 5d) models also show a reduction in training loss but with more fluctuations in validation loss, particularly in UNet, which may imply potential overfitting or sensitivity to the validation set. Among all models, NL-LinkNet-SSR achieves the lowest training and validation losses, indicating superior learning and generalization capabilities. The validation loss closely tracks the training loss with a minimal gap, suggesting robust performance with less overfitting.

Overall, NL-LinkNet-SSR exhibits the most promising results in minimizing both training and validation losses, followed by NL-LinkNet and DeepLabv3+. UNet and FCN show greater variance in validation loss, indicating potential challenges in generalization.

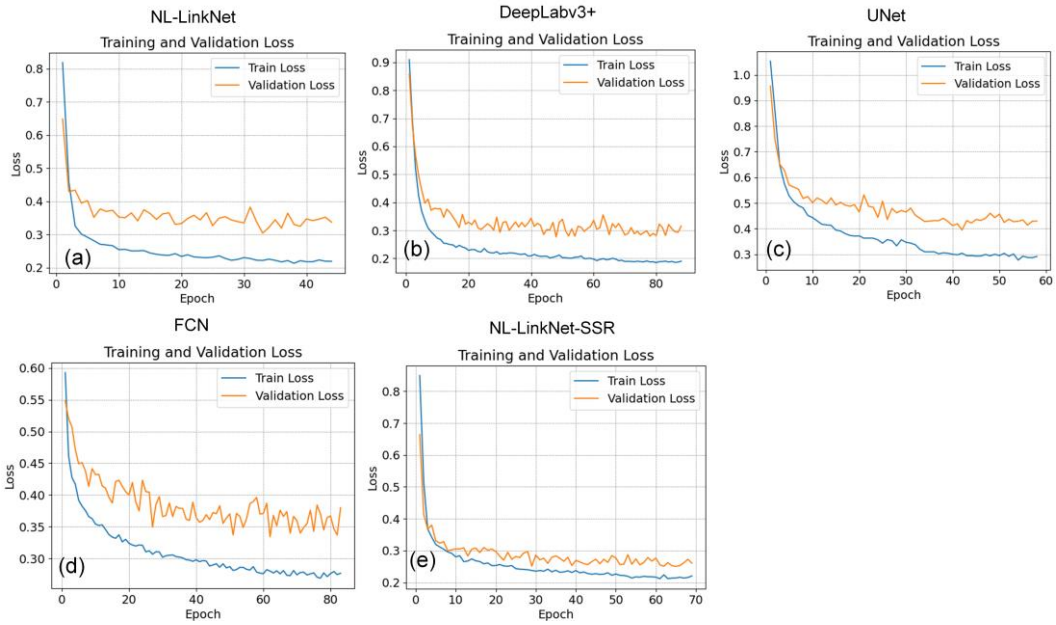

Figure 5. Training and validation loss convergence curves for different models on the railway dataset. The proposed model shows the smoothest and lowest loss values compared to the other models. (**a**) NL-LinkNet; (**b**) DeepLabv3+; (**c**) UNet; (**d**) FCN; and (**e**) NL-LinkNet-SSR.

5.2. Ablation Experiment

To validate the impact of the proposed modules and improvements on the performance of the railway track detection network, ablation experiments were designed to quantitatively evaluate track line extraction performance. The experiments were based on the NL-LinkNet network, with two enhancements individually incorporated: the SimAM attention mechanism and the Sobel edge detection module. This resulted in the NL-LinkNet-SimAM and NL-LinkNet-Sobel networks, respectively. When both modules were combined, the network was designated as NL-LinkNet-SSR. The experimental results for these networks on the test dataset are presented in Table 3.

Table 3. The evaluation metrics of the ablation experiment on the railway track extraction task.

Method	Accuracy	F1-Score	Miou	Recall	Kappa
NL-LinkNet	0.960	0.706	0.618	0.733	0.810
NLLinkNet-Sobel	0.971	0.718	0.620	0.758	0.824
NLLinkNet-SimAM	0.975	0.732	0.644	0.761	0.835
NL-LinkNet-SSR	0.978	0.749	0.653	0.780	0.841

The visual results of the ablation experiments are illustrated in Figure 6, highlighting the differences in track line extraction performance among the various models. The base NL-LinkNet model, as shown in Figure 6c, tends to miss or inaccurately detect track lines in complex scenarios. Incorporating the Sobel edge detection module, resulting in the NL-LinkNet-Sobel model, leads to noticeable improvements in the completeness and accuracy of track line extraction, especially in identifying track line edges, as depicted in Figure 6d. The NL-LinkNet-SimAM model, which includes the SimAM attention mechanism, further enhances track line detection by improving focus on track lines and reducing false detections and omissions, particularly in complex backgrounds, as seen in Figure 6e. Finally, the NL-LinkNet-SSR model, which combines both the Sobel edge detection and SimAM attention mechanisms, demonstrates the best performance in track line extraction, almost perfectly reconstructing the actual track lines with minimal false detections and omissions, as shown in Figure 6f. This indicates that the combination of edge detection and attention

mechanisms effectively enhances the model's detection capabilities in complex scenarios, significantly improving accuracy and completeness.

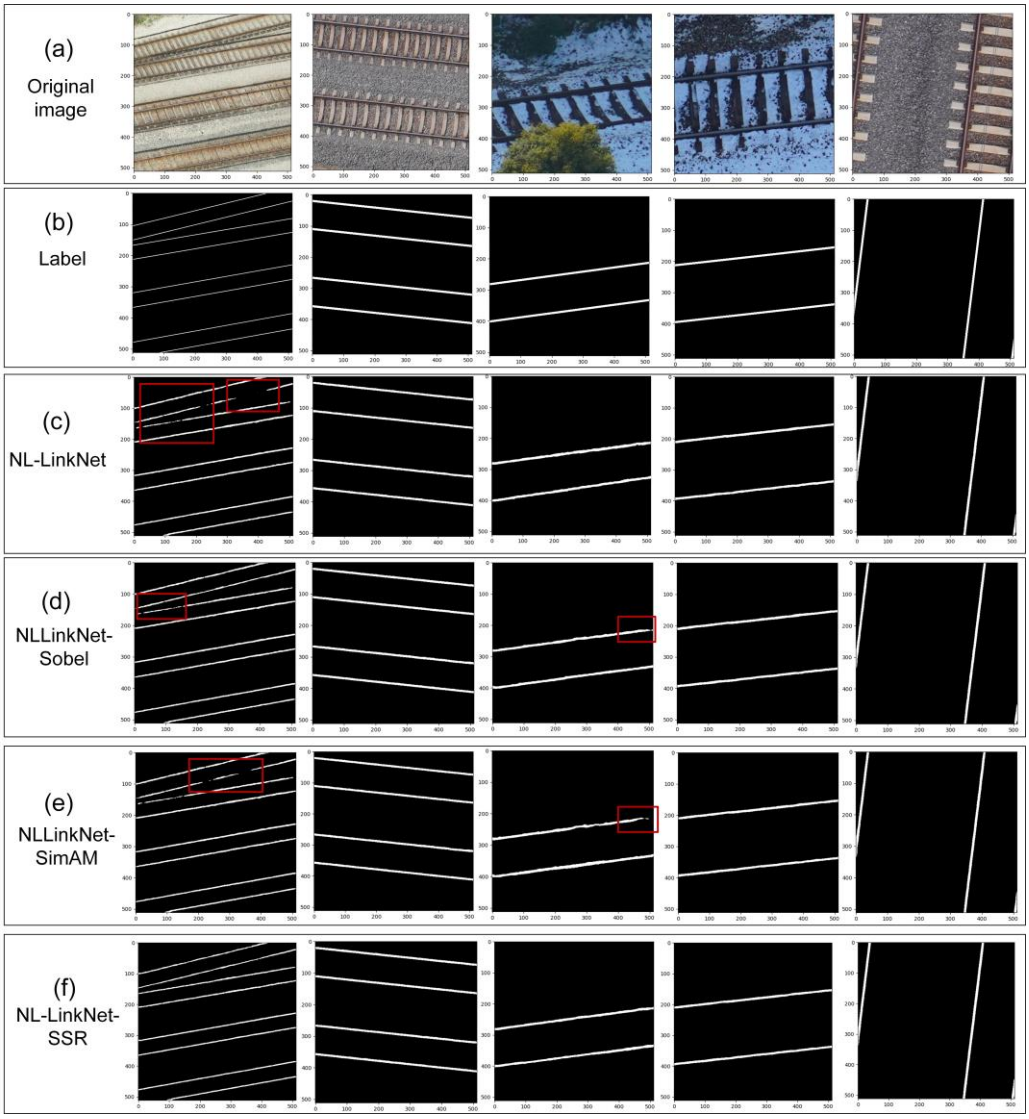

Figure 6. Ablation Study Results of Railway Track Line Extraction. (**a**) UAV imagery. (**b**) ground truth. (**c**) NL-LinkNet. (**d**) NLLinkNet-Sobel. (**e**) NLLinkNet-SimAM. (**f**) NL-LinkNet-SSR.

Table 3 supports these visual observations with quantitative results from the ablation experiments. The base NL-LinkNet model achieves an accuracy of 0.960 and an F1-Score of 0.706. Adding the Sobel edge detection module (NL-LinkNet-Sobel) improves all metrics, with accuracy increasing to 0.971 and the F1-Score rising to 0.718. The incorporation of the SimAM attention mechanism (NL-LinkNet-SimAM) further enhances performance, with accuracy reaching 0.975 and an F1-Score of 0.732. The NL-LinkNet-SSR model, integrating both modules, achieves the highest performance, with an accuracy of 0.978 and an F1-Score of 0.749. These results highlight that combining both the Sobel edge detection and SimAM attention mechanisms significantly enhances the model's detection capabilities.

From the experimental results in Table 3 and Figure 6, it is evident that the NLLinkNet-Sobel, NLLinkNet-SimAM, and NL-LinkNet-SSR models show significant improvements across various metrics compared to the original NL-LinkNet model. Specifically, both the SimAM attention mechanism and the Sobel edge detection module enhance prediction accuracy to varying degrees.

The SimAM attention mechanism adaptively highlights critical features, broadens the perceptual field, and enhances responses in the target areas. This improvement leads to a reduction in missed detections and enhances the model's accuracy and robustness. In contrast, the Sobel edge detection module enriches the image's detailed features, minimizes background interference, and emphasizes edge cues, thereby improving feature extraction. This allows the model to better understand image structures and ultimately enhances prediction accuracy.

Figure 7 presents the training and validation loss curves for different models. In Figure 7a, the original NL-LinkNet model shows a rapid decrease in training loss during the early stages, but the validation loss stabilizes and fluctuates significantly in the later stages, indicating some degree of overfitting. In contrast, Figure 7b reveals that the NLLinkNet-SimAM model achieves a significantly lower and more stable validation loss compared to NL-LinkNet, suggesting that the SimAM attention mechanism effectively enhances the model's generalization ability. Similarly, Figure 7c shows that the NLLinkNet-Sobel model demonstrates improvements in both training and validation loss, further verifying the performance enhancement provided by the Sobel edge detection module. Finally, Figure 7d illustrates that the NLLinkNet-SimAM-Sobel model performs the best among all models, with the lowest training and validation losses. This indicates that the combination of the SimAM attention mechanism and the Sobel edge detection module significantly enhances the model's overall performance.

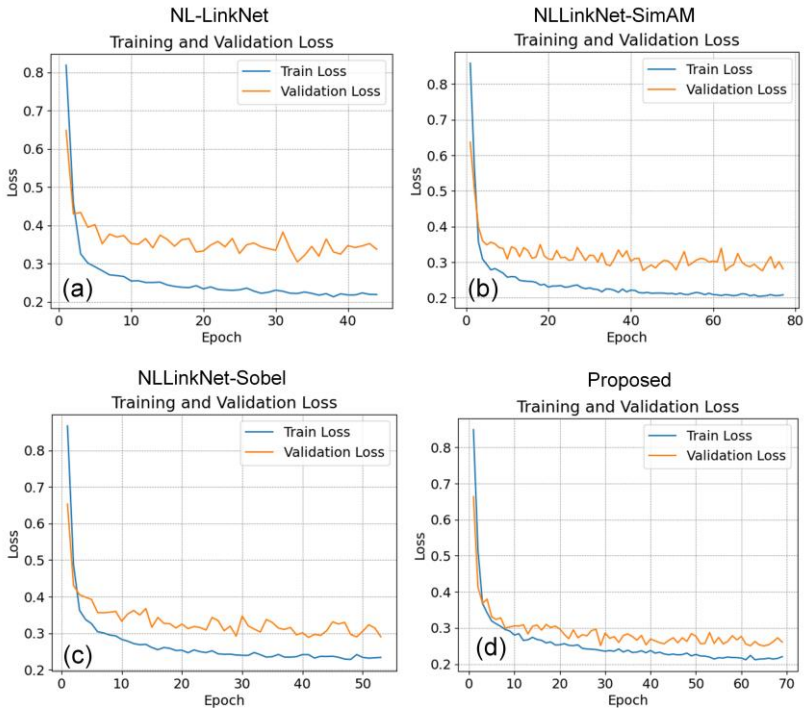

Figure 7. Training and validation loss curves for different models: (**a**) Original NL-LinkNet, (**b**) NLLinkNet-SimAM, (**c**) NLLinkNet-Sobel, and (**d**) NL-LinkNet-SSR. The figures illustrate the impact of different modules on the model's training and validation loss convergence, highlighting the benefits of incorporating SimAM and Sobel modules.

Overall, the experimental results show that the improved NLLinkNet model, integrating both the SimAM attention mechanism and the Sobel edge detection module, outperforms all other models on key metrics, offering a more reliable solution for railway track detection.

6. Discussion

6.1. Enhancements and Limitations of the Improved NL-LinkNet Model

Through the Sobel edge detection module, the improved NL-LinkNet model more accurately captures the edge features of the tracks, significantly enhancing its ability to extract track edges. This allows the model to better identify and segment track lines, even in complex backgrounds. Meanwhile, the SimAM attention mechanism adaptively emphasizes key features and expands the perceptual range, enhancing the responsiveness of target areas. This reduces missed detections and improves both the accuracy and robustness of the model. In addition, the NL-LinkNet architecture was used as the baseline for the railway track detection task due to its unique capabilities in handling complex, linear structures across expansive spatial contexts. Unlike methods that focus narrowly on localized regions, NL-LinkNet integrates spatial relationships on a broader scale, essential for accurately detecting the continuous and interconnected nature of railway tracks. This model employs non-local blocks (NLBs) to capture long-range dependencies within the input data, ensuring that each spatial feature is considered in relation to the entire scene. This global contextual awareness is crucial for distinguishing railway tracks from complex backgrounds, where similar-looking features may lead to detection errors. Moreover, NL-LinkNet is specially optimized for linear and elongated structures, providing significant advantages over more generalized attentive models that may not adequately highlight features pertinent to railway detection. Compared to other commonly used models such as FCN, U-Net, and DeepLabv3, the improved NL-LinkNet model demonstrates significant advantages across several key metrics, including accuracy, F1-Score, and MIoU. The superior performance is largely due to the SimAM attention mechanism's ability to adaptively emphasize important features after each encoder and the Sobel edge detection module's role in enriching edge detail features and reducing background interference.

However, due to the complexity of the background in the railway dataset, extracting clear, continuous, and complete railway networks in complex and varying scenes remains a challenge. Additionally, despite the significant advantages of the proposed model in various complex scenarios, such as shadows and size changes, compared to the comparative models, the overall performance improvement is still moderate. There remains a gap between the model's results and the precision of manual visual interpretation. In addition, there are certain threats to validity in our approach. One key limitation is the reliance on dataset quality, which can affect model generalization. Variations in lighting and background conditions across different regions could lead to discrepancies in performance. Another concern is the potential overfitting of the model to specific railway track patterns, although various data augmentation techniques have been applied to increase diversity, such as adding Gaussian noise, elastic transformations, and random flipping. The Sobel edge detection module is specifically designed for linear structures. The model's dependence on detecting linear features may hinder its performance when encountering more complex and non-linear railway track geometries. Future work could focus on expanding the dataset and introducing further diversity in track shapes to improve generalization.

6.2. Comparative Analysis of Improved NL-LinkNet Model with Existing Methods

The improved NL-LinkNet model proposed in this paper introduces the Sobel edge detection module and the SimAM attention mechanism, effectively enhancing railway track detection performance. Compared to existing remote sensing railway track extraction methods, such as the improved DeepLabV3+ model, which focuses on utilizing MobileNetV3 and CARAFE for efficient up sampling and accurate segmentation [31], the proposed approach emphasizes edge feature extraction to address the complexity of railway track detection. While DeepLabV3+ optimizes overall track area segmentation, our model specifically targets precise track line detection, ensuring higher accuracy in capturing track edges. In contrast to Weng et al.'s work, which uses an improved D-LinkNet model for detecting railway track areas [21], this paper focuses on extracting finer railway track line details. Weng's approach is effective at segmenting broader track regions but may miss critical edge

details, particularly in complex backgrounds. The integration of the Sobel edge detection module ensures that fine edge features are captured, improving performance in scenarios where accurate line extraction is crucial. ARTNet, with its dual-branch architecture, provides full-angle detection [20], but the NL-LinkNet-SSR model focuses on robustness in edge feature extraction, yielding higher accuracy and stability when faced with background noise and complex track shapes. Compared to the anchor-adaptive ARTNet, which is designed to handle varying railway track angles through its dual-branch architecture, the improved NL-LinkNet model demonstrates stronger global feature representation by integrating NLBs. This allows each spatial feature point to reference all other contextual information, enhancing the model's ability to detect tracks in complex backgrounds where track edges and textures may be similar to surrounding elements. Additionally, compared to "Rethinking attentive object detection via neural attention learning," the improved NL-LinkNet model offers a unique advantage by focusing on holistic scene comprehension and robust feature integration. While the neural attention learning approach focuses on dynamically prioritizing salient features within an image, our NL-LinkNet model integrates these features with non-local spatial relationships, providing a comprehensive understanding of the scene. This capability is particularly effective in environments where traditional attention mechanisms might overlook crucial interconnections of railway track features due to their localized nature. The SimAM attention mechanism, applied after each encoder in our model, further emphasizes important features while suppressing irrelevant background information, thus improving detection accuracy and reducing false positives. Moreover, the Sobel edge detection module enhances edge detail extraction, contributing to more precise track detection.

This combination of NLBs, attention mechanisms, and edge detection gives our model a clear advantage in handling diverse track shapes and challenging backgrounds, offering a significant improvement over existing methods that leverage attentive object detection primarily focused on localized areas.

7. Conclusions

To address the issues of missing and false detections in railway track detection using deep learning algorithms, this study proposes several improvements. A Sobel edge detection module is introduced into the NL-LinkNet semantic segmentation network to enhance edge segmentation performance. Additionally, the SimAM attention mechanism is integrated to focus on important features, further improving the prediction accuracy of the network model.

To overcome the challenge of limited railway track datasets, original UAV data was collected using low-altitude drones and manually annotated to create a comprehensive dataset. This dataset includes images of railway tracks captured under various environmental conditions, such as different seasons and weather scenarios, which enhances the model's robustness and generalization capability.

The proposed model outperforms all others in railway track line extraction, achieving top marks in F1-Score (0.982), MIoU (0.749), recall (0.653), and Kappa coefficient (0.841). It excels in accurately detecting and localizing tracks in complex and variable environments, demonstrating its effectiveness and reliability for practical applications. Compared to DeepLabv3+ and the original NL-LinkNet, the proposed model shows a 0.82% improvement in F1-Score and a 2.9% increase in MIoU over DeepLabv3+, as well as significant enhancements of 2.20% in F1-Score and 4.30% in MIoU compared to NL-LinkNet. These results highlight the model's improved precision and robustness in challenging conditions. The ablation studies reveal that the NL-LinkNet-SSR model, enhanced with the Sobel edge detection module and the SimAM attention mechanism, provides notable performance improvements in detecting railway tracks, thereby enhancing both the detection capabilities and accuracy of the NL-LinkNet model.

However, while the improved algorithm enhances track detection accuracy, the addition of edge detection modules and attention mechanisms increases computational

complexity. Despite these improvements, there remains considerable room for exploration in future research. For instance, combining image enhancement techniques with multitask learning methods could further improve model performance. Additionally, exploring one-shot learning and data augmentation strategies could enhance the model's effectiveness on small sample datasets. Future research could also focus on designing lightweight network models suitable for deployment on embedded devices or mobile platforms, thereby expanding the range of practical applications. Furthermore, to address the challenges encountered with our current methodologies, future efforts will focus on refining the model and employing oblique photogrammetry to capture multi-angle UAV railway track images.

Author Contributions: Conceptualization, X.F., Y.Z., and J.W.; methodology, J.W. and X.Z.; software, J.W., Z.Z., and X.Z.; validation, J.W., W.N., and Y.Q.; formal analysis, W.N., Y.Q., and N.Z.; investigation, N.Z.; resources, X.F. and Y.Z.; data curation, Z.Z.; writing—original draft preparation, J.W.; writing—review and editing, X.F.; visualization, Z.Z.; supervision, X.F. and Y.Z.; project administration, X.F.; funding acquisition, X.F. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China (Grant No. 4207133), the National Nonprofit Fundamental Research Grant of China, the China Earthquake Administration Institute of Geology (Grant Nos. IGCEA2441 and IGCEA2106), the Science and Technology Research and Development Project of China State Railway Group Co., Ltd. (Grant No. Q2023T004), and the National Key Research and Development Program of China (Grant No. 2022YFC3003700).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to restrictions related to privacy. The dataset belongs to the China Railway Design Corporation, and access is restricted to comply with the organization's data-sharing policies.

Acknowledgments: Great thanks to editors and reviewers for their constructive suggestions and insightful comments. We would like to extend our sincere gratitude to the China Railway Design Corporation for providing the UAV data that was essential for this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, J.; Cai, B.-G.; Wang, J.; Tang, T. Research on Algorithm of Electronic Track Map Data Reduction for Train Locating. *J. China Railw. Soc.* **2011**, *33*, 73–79.
2. Gibert, X.; Patel, V.M.; Chellappa, R. Deep Multitask Learning for Railway Track Inspection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 153–164. [CrossRef]
3. Zhong, J.; Liu, Z.; Yang, C.; Wang, H.; Gao, S.; Núñez, A. Adversarial Reconstruction Based on Tighter Oriented Localization for Catenary Insulator Defect Detection in High-Speed Railways. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 1109–1120. [CrossRef]
4. Wang, Z.; Wang, N.; Zhang, H.; Jia, L.; Qin, Y.; Zuo, Y.; Zhang, Y.; Dong, H. Segmentalized mRMR Features and Cost-Sensitive ELM with Fixed Inputs for Fault Diagnosis of High-Speed Railway Turnouts. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4975–4987. [CrossRef]
5. Guclu, E.; Aydin, I.; Akin, E. Development of Vision-Based Autonomous UAV for Railway Tracking. In Proceedings of the 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Virtual, 29–30 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 120–125.
6. Saini, A.; Singh, D. DroneRTEF: Development of a Novel Adaptive Framework for Railroad Track Extraction in Drone Images. *Pattern Anal. Appl.* **2021**, *24*, 1549–1568. [CrossRef]
7. Lindeberg, T. Scale Invariant Feature Transform. *Scholarpedia* **2012**, *7*, 10491. [CrossRef]
8. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
9. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
10. Soo, S. *Object Detection Using Haar-Cascade Classifier*; Institute of Computer Science, University of Tartu: Tartu, Estonia, 2014; Volume 2, pp. 1–12.
11. Osco, L.P.; Junior, J.M.; Ramos, A.P.M.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A Review on Deep Learning in UAV Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]

12. Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; Tian, Q. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 370–386.
13. Isaac-Medina, B.K.S.; Poyser, M.; Organisciak, D.; Willcocks, C.G.; Breckon, T.P.; Shum, H.P.H. Unmanned Aerial Vehicle Visual Detection and Tracking Using Deep Neural Networks: A Performance Benchmark. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, QC, Canada, 11–17 October 2021; pp. 1223–1232.
14. Liu, P.; Wang, Q.; Yang, G.; Li, L.; Zhang, H. Survey of Road Extraction Methods in Remote Sensing Images Based on Deep Learning. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2022**, *90*, 135–159. [CrossRef]
15. Tang, J.; Li, S.; Liu, P. A Review of Lane Detection Methods Based on Deep Learning. *Pattern Recognit.* **2021**, *111*, 107623. [CrossRef]
16. Gao, D.; Kang, Y.; Wang, Y. Faster R-CNN Railway Foreign Body Detection Algorithm Combined with Attention between Channels. In Proceedings of the International Conference on Signal Processing and Communication Technology (SPCT 2021), Tianjin, China, 24–26 December 2021; SPIE: Bellingham, WA, USA, 2022; Volume 12178, pp. 480–487.
17. Zhang, D.; Song, K.; Wang, Q.; He, Y.; Wen, X.; Yan, Y. Two Deep Learning Networks for Rail Surface Defect Inspection of Limited Samples with Line-Level Label. *IEEE Trans. Ind. Inform.* **2020**, *17*, 6731–6741. [CrossRef]
18. García-Santillán, I.D.; Montalvo, M.; Guerrero, J.M.; Pajares, G. Automatic Detection of Curved and Straight Crop Rows from Images in Maize Fields. *Biosyst. Eng.* **2017**, *156*, 61–79. [CrossRef]
19. Basso, M.; Pignaton De Freitas, E. A UAV Guidance System Using Crop Row Detection and Line Follower Algorithms. *J. Intell. Robot Syst.* **2020**, *97*, 605–621. [CrossRef]
20. Tong, L.; Jia, L.; Geng, Y.; Liu, K.; Qin, Y.; Wang, Z. Anchor-adaptive Railway Track Detection from Unmanned Aerial Vehicle Images. *Comput. Aided Civ. Eng.* **2023**, *38*, 2666–2684. [CrossRef]
21. Weng, Y.; Huang, X.; Chen, X.; He, J.; Li, Z.; Yi, H. Research on Railway Track Extraction Method Based on Edge Detection and Attention Mechanism. *IEEE Access* **2024**, *12*, 26550–26561. [CrossRef]
22. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-of-the-Art Review. *Remote Sens.* **2020**, *12*, 1444. [CrossRef]
23. Tao, J.; Chen, Z.; Sun, Z.; Guo, H.; Leng, B.; Yu, Z.; Wang, Y.; He, Z.; Lei, X.; Yang, J. SEG-Road: A Segmentation Network for Road Extraction Based on Transformer and CNN with Connectivity Structures. *Remote Sens.* **2023**, *15*, 1602. [CrossRef]
24. Qiu, L.; Yu, D.; Zhang, C.; Zhang, X. A Semantics-Geometry Framework for Road Extraction from Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6004805. [CrossRef]
25. Varia, N.; Dokania, A.; Senthilnath, J. DeepExt: A Convolution Neural Network for Road Extraction Using RGB Images Captured by UAV. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 18–21 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1890–1895.
26. Kestur, R.; Farooq, S.; Abdal, R.; Mehraj, E.; Narasipura, O.; Mudigere, M. UFCN: A Fully Convolutional Neural Network for Road Extraction in RGB Imagery Acquired by Remote Sensing from an Unmanned Aerial Vehicle. *J. Appl. Remote Sens.* **2018**, *12*, 016020. [CrossRef]
27. Zhu, Q.; Zhang, Y.; Wang, L.; Zhong, Y.; Guan, Q.; Lu, X.; Zhang, L.; Li, D. A Global Context-Aware and Batch-Independent Network for Road Extraction from VHR Satellite Imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 353–365. [CrossRef]
28. Dai, L.; Zhang, G.; Zhang, R. RADANet: Road Augmented Deformable Attention Network for Road Extraction from Complex High-Resolution Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5602213. [CrossRef]
29. Wang, Y.; Seo, J.; Jeon, T. NL-LinkNet: Toward Lighter but More Accurate Road Extraction with Nonlocal Operations. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
30. Mammeri, A.; Siddiqui, A.J.; Zhao, Y. UAV-Assisted Railway Track Segmentation Based on Convolutional Neural Networks. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.
31. Weng, Y.; Li, Z.; Chen, X.; He, J.; Liu, F.; Huang, X.; Yang, H. A Railway Track Extraction Method Based on Improved DeepLabV3+. *Electronics* **2023**, *12*, 3500. [CrossRef]
32. Tu, Z.; Wu, S.; Kang, G.; Lin, J. Real-Time Defect Detection of Track Components: Considering Class Imbalance and Subtle Difference between Classes. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5017712. [CrossRef]
33. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Online, 18–24 July 2021; PMLR: London, UK, 2021; pp. 11863–11874.
34. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
36. Zhang, Y.; Li, K.; Li, K.; Zhong, B.; Fu, Y. Residual Non-Local Attention Networks for Image Restoration. *arXiv* **2019**, arXiv:1903.10082.
37. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2017**, arXiv:1711.05101.
38. Yeung, M.; Sala, E.; Schönlieb, C.-B.; Rundo, L. Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [CrossRef]

39. Zhao, R.; Qian, B.; Zhang, X.; Li, Y.; Wei, R.; Liu, Y.; Pan, Y. Rethinking Dice Loss for Medical Image Segmentation. In Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 17–20 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 851–860.
40. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.