Railway Track Detection Based on SegNet Deep Learning

S. Khrueakhrai and J. Srinonchat*, *Member, IEEE*

***Abstract*— Railway track detection is crucial in railway infrastructure maintenance, safety, and operational efficiency. This paper proposes a railway track detection method based on the SegNet deep learning architecture. The SegNet model is a convolutional neural network (CNN) designed explicitly for semantic segmentation tasks. By training the SegNet model on annotated railway track images, we enable it to accurately classify each pixel in the input images as either track or non-track. The proposed method leverages the rich feature representation capabilities of deep learning to achieve robust and precise track detection, even in complex and challenging scenarios. We evaluate the performance of our approach on a benchmark dataset, considering metrics such as accuracy, intersection over union (IoU), and mean BF score. The experimental results demonstrate that our method outperforms existing track detection methods regarding accuracy and efficiency. The proposed railway track detection based on SegNet deep learning has the potential to significantly improve railway maintenance practices and enhance overall safety and operational effectiveness.**

Keywords—railway tracks detection, SegNet algorithm, deep learning.

I. INTRODUCTION

Railway track identification is essential for maintaining the effectiveness and safety of railway networks. Traditional track identification techniques sometimes rely on time-consuming, costly hand examinations or specialized equipment. However, with advances in deep learning, it is now feasible to use computer vision algorithms to locate railway tracks automatically. In order to detect railway tracks using deep learning, a deep neural network must first be trained to recognize the visual patterns and attributes connected to railroad tracks. The model may be trained to detect the distinct features of tracks, such as their form, color, and texture, by supplying a dataset of annotated pictures or video frames in which the tracks are indicated with bounding boxes. Now a day, there is a deep learning method designated specifically for identifying railway tracks. Convolutional Neural Networks (CNNs), Region-based Convolutional Neural Networks (R-CNNs), and Fully Convolutional Networks (FCNs) are a few examples of deep learning architectures and algorithms that may be used for this job. It's important to keep in mind that selecting a deep learning approach relies on the particular needs, the available data, and the nature of the task of detecting railway tracks. The applicability of different designs may be assessed in light of criteria like accuracy, efficiency, and real-time performance needs. Different architectures may have different strengths and limitations. Numerous studies have been conducted in this field, such as the investigation of the use of airborne LiDAR for track detection using deep learning networks [1], which propose a deep learning-based approach to accurately identify and locate railway tracks in LiDAR point cloud data, contributing to railway inspection and maintenance. Integrating LiDAR data and deep learning networks for railway track detection can potentially enhance railway maintenance, monitoring, and safety. It enables automated and real-time detection of track anomalies, such as track degradation, misalignments, or foreign object presence. This, in turn, can facilitate timely maintenance interventions, reduce inspection costs, and enhance overall railway operations.The potential of satellite imagery for track detection was highlighted by Li et al [2]. This research presents a method for railway track detection from high-resolution satellite images using deep learning techniques. The authors propose an approach that leverages deep neural networks to automatically identify and localize railway tracks, facilitating railway monitoring and management. A deep learning-based method for detecting railroad tracks was presented by Chen et al. [3], who also discussed the use of deep learning networks for precise track detection. In order to automatically detect railway tracks. The results demonstrate the effectiveness of the deep learning-based approach in detecting railway tracks. The proposed method achieves high accuracy and provides reliable detection results. It outperforms traditional image processing methods and shows promise for improving efficiency and accuracy in railway track detection tasks. Also, Sun et al. [4] introduced an enhanced Faster R-CNN algorithm, highlighting the effectiveness and precision of the suggested approach. The approach consists of several key steps. First, image preprocessing techniques are applied to enhance the quality and contrast of the input images. Then, a region proposal network (RPN) generates potential track proposals in the image. The proposed RPN is optimized to improve track proposal accuracy and recall rate. Moreover, focusing on real-time track detection in various scenarios, Li et al. [5] created a railway track detection and recognition approach based on an enhanced YOLO algorithm. Deep learning methods were used for track identification in [6], the deep learning based approach for railway track detection and extraction from satellite images. A deep learning model is employed for railway track detection. Specifically, a convolutional neural network (CNN) architecture is designed and trained to recognize and classify track regions within satellite images. The CNN is trained on labeled data, where railway tracks are annotated as positive examples. To handle the challenges associated with varying track appearances and complex backgrounds, it is proposed data augmentation techniques and applies strategies to balance the training data. This helps the model generalize well to different scenarios and improves detection performance. The proposed approach

*Research supported by NRIIS No. 2589514 (FFB65E0712) and National Research Council of Thailand Grant No. 1245725 (N41D640012).

J. Srinonchat is with the Department of Electronics and Telecommunication Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand, 12110. (corresponding author to provide phone: +6689-777-5038; e-mail: jakkree_s@rmutt.ac.th).

S. Khrueakhrai, is with the Department of Electronics and Telecommunication Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand, 12110. (e-mail: saifun_k@mail.rmutt.ac.th).

is evaluated using real-world satellite images, and the results demonstrate its effectiveness in accurately detecting and extracting railway tracks. The method shows promising performance in terms of precision and recall, indicating its potential for practical implementation in railway management systems. Also, Zhai et al. [7] discussed improvements to the YOLO model for increased track detection accuracy and proposed a novel YOLO-based algorithm for detecting railway tracks. Using deep learning techniques, Choi et al. [8] addressed the problem of track anomaly detection from railway inspection car video, concentrating on locating anomalies on the track. The use of deep learning techniques for precise track identification and tracking was highlighted by Gupta et al. [9], presenting a railway track detection and tracking approach. A method that utilizes deep learning techniques for railway track detection and tracking is proposed. Their work aims to develop an automated approach that can accurately identify and track railway tracks in various scenarios, contributing to railway maintenance, safety, and operational efficiency. The proposed method involves multiple steps, including data preprocessing, feature extraction, model training, and track detection. The deep learning model is trained to distinguish between track and non-track regions in the input data, enabling it to identify the precise location of tracks. The experimental setup and evaluation metrics are used to assess the proposed method's performance. In order to investigate the use of deep learning in processing laser scanner data for track detection, Sun et al. [10] created a deep learning framework for railway track inspection. Deep learning allows for the automatic analysis and interpretation of video data, enabling the system to learn and identify various track anomalies, such as cracks, deformations, missing components, or other irregularities. By leveraging the power of deep neural networks, the method can potentially provide accurate and real-time detection of anomalies that may otherwise be challenging to identify manually.

Although deep learning models can learn complex patterns and representations from large amounts of data, it requires a large amount of labeled data for training. This study employs SegNet technology to identify railroad track improvement.

II. Methodology and Related Works

A deep learning architecture called SegNet is explicitly created for semantic segmentation tasks. An encoder-decoder structure with skip links is used to conduct pixel-level picture segmentation. In order to successfully restore the spatial resolution and precisely categorize each pixel in the input picture, the model learns to encode and decode high-level characteristics. There are six critical components of the SegNet to achieve operation.

A. Encoder Network

The encoder network consists of several layers of convolutional and pooling operations. These layers gradually reduce the spatial resolution of the input image while extracting high-level features.

B. Decoder Network

The decoder network reconstructs the high-resolution output from the encoded features, and the decoder aims to restore any lost spatial information.

C. Skip Connections

An essential part of SegNet is skip connections [11-13] which connect the corresponding encoder and decoder layers. Skip connections aid in maintaining spatial data and enhancing segmentation precision.

D. Max Pooling

SegNet stores the pooling indices during the max-pooling operations in the encoder to facilitate effective upsampling. These indices show where the maximum values are located within each pooling region.

E. Softmax Classification

A softmax classification layer is used as the decoder network's top layer. Each pixel is given a class probability representing the likelihood of belonging to a particular class. This makes it possible to classify and segment the input image at the pixel level.

F. Training

SegNet is typically trained using backpropagation and optimization techniques like stochastic gradient descent (SGD) to minimize a loss function. Pixel-wise cross-entropy loss assesses the difference between the ground truth labels at each pixel and the predicted class probabilities and is a popular choice for semantic segmentation tasks.

The SegNet has been used in many works for object detection. The work [13] introduced SegNet as an encoder-decoder network with skip connections to accurately segment images at the pixel level. The encoder network consists of convolutional and pooling layers, gradually reducing the spatial resolution while extracting high-level features. The decoder network reconstructs the high-resolution output using upsampling and deconvolutional layers. It is shown the importance of preserving spatial information during the upsampling process. SegNet addresses this by storing the pooling indices during the max-pooling operations in the encoder. These indices are used during upsampling to accurately place the decoded features in their original positions, ensuring precise reconstruction.

The application [14] focuses on comprehending urban scenes and semantic SegNet segmentation. The study suggests a technique for performing instance-level semantic segmentation of urban scenes that combines SegNet with a multiple instance detection network. The method achieves precise object segmentation in intricate urban settings, revealing insights into scene comprehension and facilitating uses in autonomous driving and urban planning. In the context of autonomous driving, the application [15] focuses on using SegNet for lane detection and segmentation. In order to accurately segment lanes from RGB-D (color and depth) data, the study presents a hierarchical CNN architecture incorporating SegNet. Robust lane detection is achieved by the suggested technique, which is necessary for autonomous vehicles to understand and navigate the road environment. It shows how SegNet could improve the dependability and safety of autonomous driving systems.

III. Experiment Setup

The experiment setup for this research plays a crucial role in achieving accurate and reliable results. It involves carefully designing and configuring various experiment components, including the dataset, model architecture, hyperparameters, and evaluation metrics.

A. Dataset

A dataset of 2,000 images of railway tracks, separated into training and testing sets as 70% and 30%, respectively, was utilized in this study to train and evaluate the SegNet mode. 640x480 pixels are used in each image.

(a) (b)

(c) (d)

Figure 1. Example of original images

B. Model Architecture

The proposed SegNet architecture is designed with an encoder-decoder structure with skip connections to achieve accurate and efficient segmentation results. The max pooling and stride functions are used, with each size as 2x2.

C. Hyperparameters

The hyperparameters of the proposed model are shown in Table 1.

TABLE I. Hyperparameters

Parameter	Value
Number of Convolution Layers	26 layers
Filter Sizes	3x3
Number of Filters	64, 128, 256 and 512
Optimizer	SGDM
Momentum	0.9
BatchSize	1
Epochs	50
LearnRate	0.01, 0.001, 0.0001, 0.00001
Activation Function	Rectified Linear Unit (ReLU)
Loss Function	Cross-Entropy Loss

D. Evaluation Metrics

The Accuracy, Intersection over Union (IoU), and Mean BF-Score or F1-Score [16] are used evaluation metrics to evaluate the SegNet model.

1) Accuracy

Accuracy is a straightforward evaluation metric that measures the overall correctness of the predicted segmentation compared to the ground truth labels. It is calculated as the ratio of correctly classified pixels (true positives and true negatives) to the total number of pixels in the image, as shown in (1).

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (1)$$

Where:

TP (True Positives) is the number of correctly classified positive instances (pixels or samples).

TN (True Negatives) is the number of correctly classified negative instances.

FP (False Positives) is the number of falsely classified positive instances.

FN (False Negatives) is the number of falsely classified negative instances.

2) Intersection over Union (IoU)

Intersection over Union (IoU) is a more informative evaluation metric for semantic segmentation. IoU indicates how well the model captures the spatial extent of a particular class, as shown in (2). Higher IoU values indicate better segmentation accuracy for that class.

$$IoU = \frac{Intersection\ Area}{Union\ Area} \quad (2)$$

Where:

Intersection Area: The area of overlap between the predicted segmentation and the ground truth labels for a specific class.

Union Area: The combined area of the predicted segmentation and the ground truth labels for that class, including overlapping and non-overlapping regions.

3) The Mean BF-Score

The Mean BF-Score, or Boundary F1-Score, is a metric that evaluates the model's performance in capturing the boundaries between different classes. The Mean BF-Score is calculated by comparing the predicted boundary map with the ground truth boundary map, as shown in (3). It considers both precision and recall of the boundary detection and provides a single score that represents the model's performance in capturing object boundaries. Higher Mean BF-Score values indicate better boundary detection accuracy.

$$Mean\ BF-Score = \frac{2*(Precision*Recall)}{(Precision+Recall)} \quad (3)$$

Where:

Precision: Precision measures the fraction of correctly predicted boundary pixels (True Positives) out of all the pixels predicted as boundary (True Positives + False Positives). It represents the accuracy of boundary detection.

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

IV. Results

Table II shows that the experiment's findings revealed that different performance accuracy depended on the learning pace.

TABLE II. Evaluation of the SegNet model

Evaluation of the SegNet model			
Learning Rate	Accuracy	IoU	Mean BF-Score
0.01	0.9799	0.9593	0.9613
0.001	0.9688	0.9433	0.9568
0.0001	0.9574	0.9158	0.9097
0.00001	0.8602	0.7503	0.5337

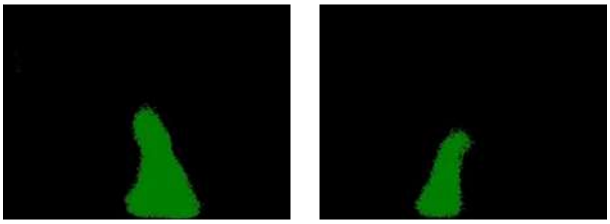
Figures 2 and 3 show the categorization of railroad tracks. It is evident that the value of SegNet's learning rate directly impacts the increase of images' quality.

(a) Original image

(b) LR=0.01

(c) LR=0.001

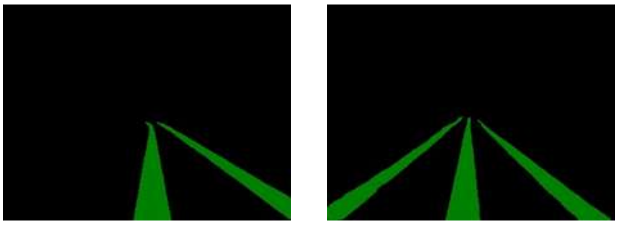
(d) LR=0.0001



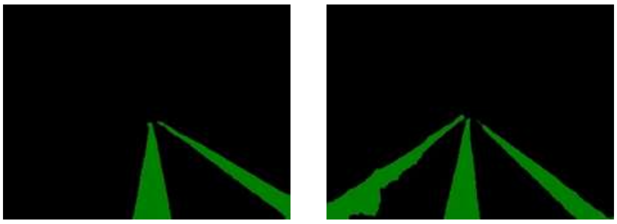
(e) LR=0.00001

Figure 2. Single railway tracks segmentation

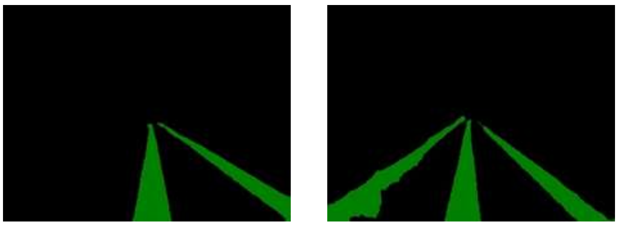
(a) Original image



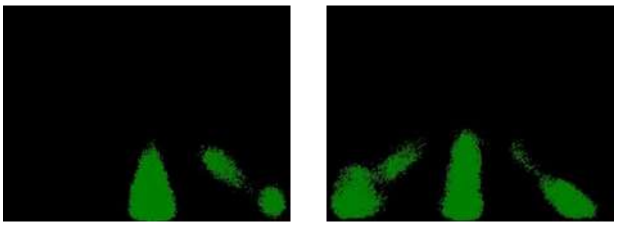
(b) LR=0.01



(c) LR=0.001



(d) LR=0.0001



(e) LR=0.00001

Figure 3. Multi-railway tracks segmentation

V. CONCLUSION

This study shows how to recognize railway tracks using deep learning based on SegNet. The suggested system's input is made up of 2,000 images of railway tracks, which are divided into training and testing segments with corresponding weights of 70% and 30%. The stride function and maximum pooling are both set to 2x2. As illustrated in Table 1, the SegNet structure is configured with a variety of options. It is clear that the suggested system performed best at 0.9799, 0.9593, and 0.9613 using three assessment methodologies under the learning condition of 0.01. Future study will concentrate on adjusting the SegNet structure to increase accuracy performance in the learning rate and also look into hybrid deep learning.

ACKNOWLEDGMENT

This study was supported by NRIIS No. 2589514 (FFB65E0712) and National Research Council of Thailand Grant No. 1245725 (N41D640012). Additionally, we would like to extend our heartfelt appreciation to the Signal Processing Research Laboratory at Rajamangala University of Technology Thanyaburi, Faculty of Engineering for their knowledge and experience.

REFERENCES

[1] Y. Li, Y. Xia, Z. Wang, Z. X u and X. Li, "Railway Detection from Airborne LiDAR Data Using Deep Learning Networks," Remote Sensing, 11(2), pp.184, 2019.

[2] Y. Li, Y. Wu and Y. Xia, "Railway track detection from high-resolution satellite images using deep learning," Remote Sensing, 12(17), pp. 2776, 2020.

[3] L. Chen, H. Wang, J. Wang, and Z. Wei, "A Railway Track Detection Method Based on Deep Learning," In 2020 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), pp. 156-159, 2020.

[4] L. Sun, W. Wang, Y. Lu, Z. Hu nad B. Shang, "An automatic approach to railway track detection using improved Faster R-CNN," IEEE Access, 7, pp. 81994-82005, 2019.

[5] W. Li, X. Xie, J. Xie and Y. Wu, "Railway Track Detection and Recognition Based on Improved YOLO Algorithm," Proceedings of the 4th International Conference on Computer Science and Application Engineering, pp. 123-127, 2019.

[6] D. Sharma, S. Banerjee and R. Khandelwal, "Deep Learning Based Railway Track Detection and Extraction Using Satellite Images," International Conference on Computing, Power and Communication Technologies, pp. 85-91, 2021 Springer.

[7] Z. Zhai, B. Shen and X. He, "A Novel Railway Track Detection Method Based on Improved YOLO Algorithm," International Conference on Information Science and Communications Technologies, pp. 119-126, 2021.

[8] J. Y. Choi, D. Yoon, S. Lee, and J. H. Kim, "Deep learning-based railroad track anomaly detection from railway inspection car video. Applied Sciences, 10(15), 5256, 2020.

[9] R. Gupta, D. B. Rawat and R. P. Maheshwari, "Railway Track Detection and Tracking using Deep Learning Techniques," IEEE International Conference on System, Computation, Automation and Networking, arXiv:1511.00561v3, pp. 1-6, 2021.

[10] Z. Sun, S. Wu, J. Yang and B. Chen, "A Deep Learning Framework for Railway Track Inspection Using Laser Scanner Data," Remote Sensing, 10(9), 1437, 2018.

[11] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury and C. Pal, "The Importance of Skip Connections in Biomedical Image Segmentation," International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 179-187, 2016, DOI:10.1007/978-3-319-46976-8-19

[12] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), pp. 2481-2495, 2017, DOI: 10.1109/TPAMI.2016.2644615.

[13] F. Isensee, J. Petersen, A. Klein, et al., "NNU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation," IEEE Transactions on Medical Imaging, 40(2), pp. 525-536, 2021, DOI: 10.1109/TMI.2020.3019565.

[14] M. Siam, M. Gamal, M. Elhoseiny and A. Elgammal, "Multiple Instance Detection Network with Online Instance Classifier Refinement," Proceedings of the European Conference on Computer Vision (ECCV), pp. 374-391, 2018.

[15] X. Pan, S. You, X. Wang and H. Lu, "Hierarchical CNN with Room-Attention for RGB-D Semantic Segmentation," IEEE Transactions on Intelligent Transportation Systems, 20(6), 2024-2034, 2018.

[16] A. Prommakhot and J. Srinonchat, "Scaled Dilation of DropBlock Optimization in Convolutional Neural Networkfor Fungus Classification," Computers, Materials & Continua, 72(2), 3313-3329, 2022.