Transformer-based Pavement Crack Tracking with Neural-PID Controller on Vision-guided Robot

Jianqi Zhang[1,2], Xu Yang[2,3,*], Wei Wang[1], Ioannis Brilakis[4], Hainian Wang[3] and Ling Ding[5]

[1]School of Information Engineering, Chang'an University, Xi'an, 710064, China
[2]School of Future Transportation, Chang'an University, Xi'an, 710064, China
[3]School of Highway, Chang'an University, Xi'an, 710064, China
[4]Laing O'Rourke Centre, Engineering Department, Cambridge University, Cambridge, CB2 1PZ, United Kingdom
[5]School of Transportation Engineering, Chang'an University, Xi'an, 710064, China

jqzhang@chd.edu.cn, yang.xu@chd.edu.cn, wei.wang@chd.edu.cn, ib340@cam.ac.uk, wanghn@chd.edu.cn

Abstract -

Pavement crack tracking in unstructured road environments has been and continues to be a crucial and challenging task, playing a vital role in achieving accurate crack sealing for automated pavement crack repair. However, slender cracks suffer from insufficient feature extraction and low tracking efficiency. In this article, a hybrid adaptive control scheme combined with a self-tuning neural network and proportional–integral–derivative (PID) is proposed for dynamic visual tracking of pavement cracks. Specifically, the scheme extracts crack features on the road image plane based on a S2TNet system and determines an optimal control input to guide the robot. S2TNet cross-integrates the global features through the multi-head attention module. It also adaptively recalibrates the channel responses of partial feature maps for fusion operations with the transformer module. Moreover, the Neural–PID controller is designed for adaptive adjustment of control parameters, and the scheme was validated on a physical robot platform. Extensive experimental results showed that the effectiveness of the proposed method in achieving real-time tracking for pavement cracks.

Keywords -

Crack Tracking; Crack Segmentation; Transformer; Neural–PID Control; Mobile Robot

1 Introduction

Pavement cracks are prevalent and hazardous defects that significantly impact driving safety in highway transportation. They primarily arise from a range of factors, such as heavy traffic loads, subpar construction practices, the influence of climate, and inadequate drainage[1, 2]. Failure to promptly repair pavement cracks can lead to accelerated deterioration of the pavement structure through the ingress of rainwater. Even a small crack can rapidly degrade into a pothole overnight, posing a significant hazard to high-speed driving[3, 4]. Hence, regular maintenance and repair of pavement cracks are imperative to prevent crack deterioration and ensure traffic safety[5, 6]. Manual sealing is the conventional approach for repairing pavement cracks. However, manual pavement repair proves to be time-consuming, expensive, and subjective. Therefore, there is a growing demand for automated and efficient repair methods in pavement crack tracking.

Recent studies have primarily focused on the development of crack segmentation with convolutional neural network (CNN)-based methods in road environments. For instance, [7] constructed a novel crack segmentation network called CrackW-Net, and designed the skip-level round-trip sampling block, which can be easily used in various network structures. [8] developed mobile robot system can effectively segment pavement cracks in real scenarios at a speed of 25 frames per second. [9] used a 3D printer as a crack-filling machine. In recent years, path tracking research based on mobile device motion control has become popular. A crack sealing system was designed to control the experimental three-dimensional (3D) printer to repair cracks[2]. [10] proposed the cross-entropy-based adaptive fuzzy control for crack tracking with VT-UMbot.

The insufficient feature extraction is significantly triggered by the limited receptive field in the CNN segmentation model and it often leads to a coarse segmentation of the cracks. Over the years, researchers have proposed various techniques to improve object detectability. These approaches include encoder-decoder[11], multi-scale attention[12], and multi-scale feature extraction[13]. Additionally, efforts have been made to enhance object feature representation[14] and fusion[15]. However, despite these advancements, challenges still persist in the field, such as inadequate detection of detailed features and susceptibility to background lighting conditions. On the other hand, low tracking efficiency is also caused by Slender pavement cracks have extreme length-width ratio and complex topology, which lead to irregular paths. Path tracking research mainly focuses on distribution rules and trajectory obeying certain rules. Recent tracking control methods range from traditional PID to various optimized and improved PID such as fuzzy control[10], genetic algorithms[16] and ant colony algorithms[17]. However, challenges related to tuning of control parameters in specialized environments significantly impact the performance of path tracking.

This article presents a pavement crack tracking framework that enhances tracking efficiency in unstructured road scenarios by fusing real-time crack video context features with transformer-based segmentation and proposing Neural–PID control strategies in the crack tracking. To address the insufficient feature extraction and low tracking efficiency, extensive experiments are conducted and verified. The contributions of this work are fourfold:

- Aiming at the problem of pavement crack tracking, a joint transformer-based fusion model and Neural–PID tracking control scheme is proposed. This algorithm successfully achieves stable real-time tracking for pavement crack.
- Enhancing the performance and effectiveness of crack segmentation in challenging road conditions with insufficient feature extraction. This article Introduces a transformer-based fusion model, which leverages multi-fusion strategies to address the challenges posed by coarse crack feature extraction.
- Considering pavement cracks with slender shape and irregular path, a Neural–PID tracking control method is proposed to improve the performance of tracking. Specifically, adaptive adjustment of control parameters is achieved by neural network.
- Conducting extensive experiments on self-created S2T-Crack dataset, the proposed algorithm is successfully deployed in self-developed vision-guided robot. The results show that our method achieved State-of-The-Art.

The structure of this article is organized as follows. Section 2 provides the existing related work. Section 3 outlines the detailed design of our methodology. Section 4 presents the experimental validation of our approach. Finally, Section 5 summarizes the article and discusses future directions.

2 Related works

This section reviews the literature relevant to our proposed pavement crack tracking.

Crack Segmentation. Crack segmentation is a crucial distress inspection technique for different infrastructures, including roads, bridges, tunnels, airports and buildings. There are numerous crack segmentation methods developed based on deep learning. YOLOv5[18] is a single-stage object detection model known for its architectural features such as the incorporation of Cross-Stage Partial (CSP) and Spatial Pyramid Pooling-Fast (SPPF) methods in the backbone network, as well as the utilization of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) in the Neck network. A lightweight pavement crack detection model is proposed to realize the dual tasks of object detection and semantic segmentation[19].

However, CNN models primarily focus on local feature extraction, which may result in information ambiguity and coarse segmentation when dealing with long-range dependency relationships. Therefore, this research aims to fuse YOLOv5 with Transformer to achieve effective crack segmentation.

Vision Transformer. Thanks to strong representation capabilities, researchers are looking at ways to apply transformer to computer vision tasks. In various visual benchmarks, the performance of the transformer-based model is similar to or better than other CNN types of networks. [20] classified these visual transformer models according to different tasks, and analyzes their advantages and disadvantages, so as to review them. A new video instance segmentation framework based on Transformer is proposed, called VisTR, which regards the VIS task as a direct end-to-end parallel sequence decoding / prediction problem[21]. [22] designed a segmentation model called SEgmentation TRansformer (SETR). A large number of experiments show that SETR has achieved competitive results on Cityscapes.

Compared to CNN, transformer incurs higher computational costs and longer training times. Given the subtle nature of crack features, achieving fine-grained segmentation of cracks is crucial. Therefore, this research introduces self-attention and cross-attention mechanisms to enhance feature extraction.

PID Control. PID control is widely used in path tracking control of mobile robots. In the absence of robot knowledge, the PID controller may be the best controller because it is model-free and its parameters can be easily adjusted separately. However, the parameters depend on artificial empirical values, and parameter optimization is an existing challenge. [23] used the adaptive PID controller to adjust the error to adjust the front wheel angle. A robust PID controller for flight control of four-rotor aircraft is proposed[24]. An adaptive fuzzy control (CEAFC) method based on cross entropy is proposed for PID parameter tuning[10].

Traditional PID controllers are susceptible to external disturbances when it comes to parameter adjustments, leading to convergence issues and system uncertainty. To address these challenges, this study proposes the Neural-PID approach to ensure effective tracking performance.

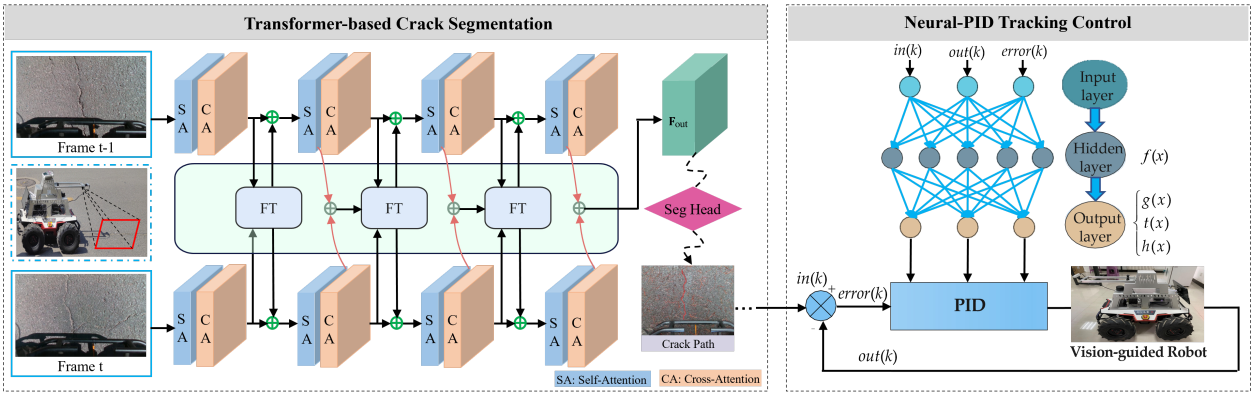

Figure 1. General framework of our proposed scheme for pavement crack tracking on vision-guided robot. It mainly includes two separate modules: transformer-based crack segmentation (including two branches and three fusion modules), Neural-PID crack tracking (containing three layers networks). All modules are implemented based on the unified YOLOv5 framework, and the details of each module are shown in Figure 2. It is worth noting that both the input video images and tested results were conducted on the S2TCrack dataset.

3 Methodology

This work first describes related issues of pavement crack tracking systems. Additionally, it is deployed on vision-guided robot to achieve crack tracking. This section presents the details of our proposed method.

3.1 Framework

This article focuses on two key aspects of crack tracking in road environment. Firstly, it addresses the challenge of achieving accurate crack segmentation in pavement scenarios characterized by slender crack and complex background. Secondly, it examines the low tracking efficiency of crack tracking control methods in limited parameters tuning conditions. To address these challenges, a crack tracking framework is proposed that ensembles transformer-based fusion network and Neural–PID tracking control algorithm. This framework, illustrated in Figure 1, comprises two main modules: transformer-based crack segmentation and Neural–PID tracking control. The feature fusion module employs the yolov5 under the popular transformer to encode and decode crack video images, enabling the fusion of image pixels at the feature level. In order to adaptively tune the tracking controller parameters more quickly, a three-layer structured neural network is used. A detailed overview of the framework is presented in the subsequent subsections.

3.2 Crack Segmentation with Transformer

The proposed module employs the yolov5 under the popular transformer to encode and decode crack video images, enabling the fusion of image pixels at the feature level. In contrast to the initial iteration of YOLOv5, this study presents a novel approach that incorporates a two-branch convolutional neural network backbone. This backbone is illustrated by the light-green modules in Figure 2, and it is designed to extract crack features between video frames from a vision-guided robot. In the context of fusion utilizing FT modules, the fusion process occurs at three distinct stages, facilitating the integration of fused characteristics that comprise both coarse-grained and fine-grained semantic information.

A common layer in the encoder and decoder structure is multi-head attention, which consists of multiple parallel self-attention mechanisms. In Self-Attention, Q, K, and V are three vectors calculated on the same input (such as a word in a sequence). Specifically, Q, K, and V can be obtained by applying a linear transformation (e.g., using a fully connected layer) to the original input word's embedding. The dimensions of these three vectors are usually the same and depend on the decisions made during the model design. During the computation of Self-Attention, Q, K, and V are used to calculate attention scores, representing the relationship between the current position and other positions. Attention scores are obtained by taking the dot product of Q and K, dividing the scores by 8, and applying softmax normalization. This process yields weights for each position. Next, these weights are used to compute the weighted sum of V, resulting in the output for the current position. In order to illustrate the effectiveness of our proposed FT fusion module, the feature extraction network of YOLOv5 is extended and redesigned as a backbone composed of two streams to achieve modal fusion and interaction.

3.3 Neural–PID Control for Crack Tracking

In the process of actual pavement crack path tracking motion control, due to the complex control environment and the nonlinear and time-varying characteristics of the controlled object, the conventional PID control can not adjust the adaptive parameters and achieve good adaptability. Using the error back propagation technology, the multi-layer feedforward neural network is called to become a back propagation neural network. Because of its properties, it has excellent performance in nonlinear mapping, such as function approximation and pattern recognition. There are three layers in the back propagation neural network model: input layer, hidden layer and output layer. The input layer processes the type and quantity of input. By controlling the number of layers and activation functions, the hidden layer introduces the possibility of nonlinear mapping. The output layer is responsible for generating some information. The output of the neuron model structure is usually expressed as a nonlinear combination of input and weight.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{1}$$

The three non-negative gain parameters of the PID control scheme are output by the BP neural network, so the sigmoid function and other functions without negative output values are applied.

$$g(x) = u \cdot \frac{1}{1 + e^{-x}} \tag{2}$$

$$h(x) = \min(\max(0, x), u) \tag{3}$$

$$t(x) = u \cdot \frac{e^x}{e^x + e^{-x}} \tag{4}$$

where u is upper bound of the output. It is used to regulate the output range.

Back propagation neural network nonlinearly maps the input, output and error to the three parameters kp, ki and kd of the PID controller. In addition, the BP neural network has three neuron points for the input layer, five neuron points for the buried layer, and three neuron points for the output layer. The commonly used Tanh function is used in the hidden layer. Combined with BPNN and PID control algorithm, the online self-tuning of PID control parameters can be realized, and the optimal pavement crack tracking motion control effect can be achieved. The structure of the Neural–PID scheme is shown in Fig.1.

4 Experiments

This section focuses on evaluating the proposed method through representative benchmarks and validation. The first aspect covers the experimental settings. Then, the crack segmentation results are analyzed and discussed. Subsequently, our Neural–PID method is deployed on a vision-guided robot to achieve real-time tracking of pavement cracks.

4.1 Experimental Setting

The model training experiments were conducted on an Intel(R) i9-13900K(F) CPU running at 5.8 GHz, along with an NVIDIA GeForce RTX4090 GPU (24 GB) and the following software versions: CUDA v10.2, cuDNN v8.0.1, Pytorch v2.0, and Python v3.8. The unmanned wheeled robot is equipped with an embedded Nvidia Jetson AGX Xavier computer, serving as the main processor with the following specifications: 512 CUDA cores and 64 tensor cores within an Nvidia Volta GPU, v8.2 ARM CPU with 8 cores, and 32 GB DDR4 memory. To acquire pavement crack video images in the front view scene of the unmanned wheeled robot, a front-mounted Realsense D435i camera with a 135-degree field of view (FOV) and an RGB-D perception unit is utilized. The embedded environment includes Jetpack 4.4, PyTorch 1.8, Linux Ubuntu 18.04, and ROS Melodic, as shown in Figure 3.

The evaluation metrics utilized to assess the performance of our proposed method are Precision (B), Precision (M), Recall (B), Recall (M), and AP (Average Precision). Furthermore, the AP incorporates mAP0.5 (B), and mAP0.5 (M), which represent the AP with an IoU threshold greater than 0.5, and mAP0.5:0.95 (B), mAP0.5:0.95 (M), which pertain to the average AP with an IoU threshold ranging from 0.5 to 0.95 in increments of 0.05. The

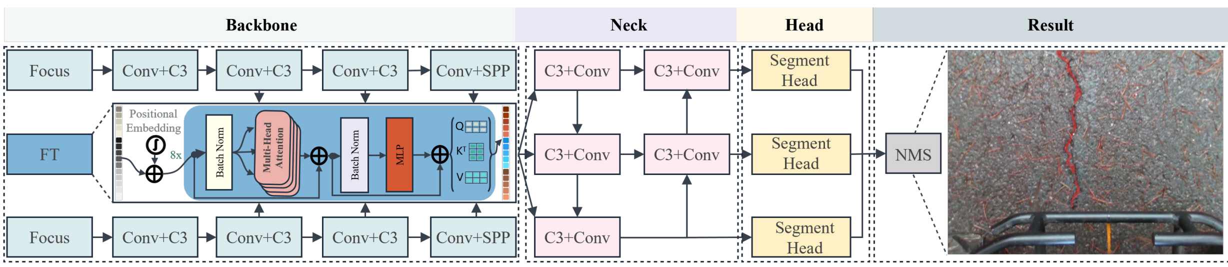

Figure 2. The architecture of YOLOv5 uses a fusion transformer method that encompasses four separate components: backbone, neck, head, result.

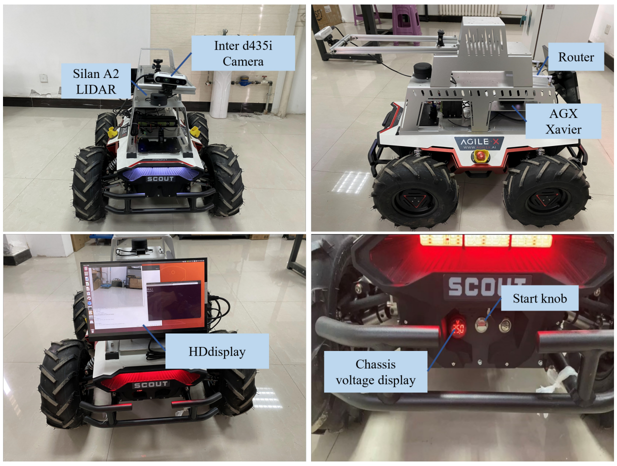

Figure 3. Working conditions of our vision-guided robot under different perspectives are displayed.

notation (B) represents the metric of the predicted boundary frame, corresponding to crack detection. Similarly, the notation (M) represents the metric of the binary mask, corresponding to crack segmentation.

4.2 Results of Crack Segmentation

This section presents an approach to significantly enhance the performance of crack segmentation using the proposed method. The experimental results are analyzed on the open data set CFD and the self-built data set S2TCrack.

4.2.1 CFD Dataset

CFD is utilized for evaluation. The CFD dataset comprises 118 pavement crack images, each with dimensions of 480 pixels by 320 pixels. These images were captured by individuals standing on the road using an iPhone. The ground truths were meticulously annotated at the pixel level, a task that demands significant labor. The images exhibit high quality with a smooth and clean background. Table 1 compares the performance of YOLOv5, our method (Ours), on the pre-trained models n, s, m, l, x. Our method, using the different pretrained model, demonstrated improved performance on the CFD dataset. The following best performance metrics are: [Precision(M) = 0.6818, Recall(M) = 0.5178, APval0.5(M)=0.5304, APval0.5:0.95(M)=0.2453]. Moreover, based on the comprehensive results obtained from the CFD dataset, our proposed method exhibits significantly better performance and versatility, showcasing its exceptional ability in pixel-level crack segmentation tasks.

4.2.2 S2T-Crack Dataset

This section also includes a comparative experiment on the self-built S2TCrack dataset, as presented in Figure4. Our method demonstrates superior segmentation performance in the pretrained model 's', which boasts a mere 6.7M parameters and 15.2M GFLOPs. Meanwhile, the segmentation accuracy is moderately acceptable. Along with the segmentation results of three scenes from the self-built S2TCrack dataset, YOLOv5 roughly splits the cracks, ignoring certain subtle features, which may result in incomplete masks, leading to fractures or local losses. Our method effectively generates masks that appropriately cover the target cracks, thanks to the utilization of SA and CA. To further enhance the performance, FT modules are integrated to fuse crack features. Our method is capable of generating highly accurate binary masks, making it suitable for various complex scenes.

4.3 Online Tuning of PID Parameters

This section presents an approach to significantly enhance the performance of crack tracking using the proposed method. The experimental results are analyzed on different control algorithms.

4.3.1 Comparison of Tracking Control

As shown in Figure 5, compared with CEAFC, the Neural-PID control scheme approaches the ideal solution with a faster convergence rate at iteration 200, indicating that the Neural-PID has stronger deterministic global search ability and faster high-dimensional optimal solution discovery speed. The results show that the Neural-PID control algorithm is superior to the other three methods.

Table 1. Real-time segmentation results in the CFD dataset.

Method	Pretrained Model	Batch_Size	Precision(B)	Precision(M)	Recall(B)	Recall(M)	$mAP^{val}0.5(B)$	$mAP^{val}0.5(M)$	$mAP^{val}0.5:0.95(B)$	$mAP^{val}0.5:0.95(M)$	Params /M	GFLOPs /M
YOLOv5	n	32	0.6688	0.4424	0.4523	0.3158	0.4339	0.2301	0.1644	0.0393	1.9	6.7
	s	16	0.7254	0.4621	0.4474	0.4474	0.4944	0.3875	0.2456	0.0854	7.4	25.7
	m	8	0.7326	0.4562	0.4645	0.3947	0.4750	0.3631	0.2586	0.0545	21.7	69.8
	l	2	0.7289	0.4637	0.4737	0.4211	0.5032	0.3849	0.2561	0.0911	47.3	146.4
	x	2	0.7288	0.4726	0.5256	0.4211	0.4865	0.3756	0.2871	0.0822	88.2	264
Ours	n	32	0.7153	0.5958	0.5037	0.4167	0.5723	0.4548	0.3264	0.1967	2.0	6.9
	s	16	**0.7982**	0.5653	0.4943	0.4817	**0.5921**	0.4906	0.3485	0.1873	7.5	25.7
	m	8	0.7705	0.5831	0.5736	0.4524	0.5257	0.5187	0.3356	0.2294	21.8	69.9
	l	2	0.7657	0.6024	**0.5975**	0.4688	0.5354	0.5018	0.3721	**0.2453**	47.4	146.7
	x	2	0.7724	**0.6818**	0.5487	**0.5178**	0.5677	**0.5304**	**0.3953**	0.2102	88.4	265

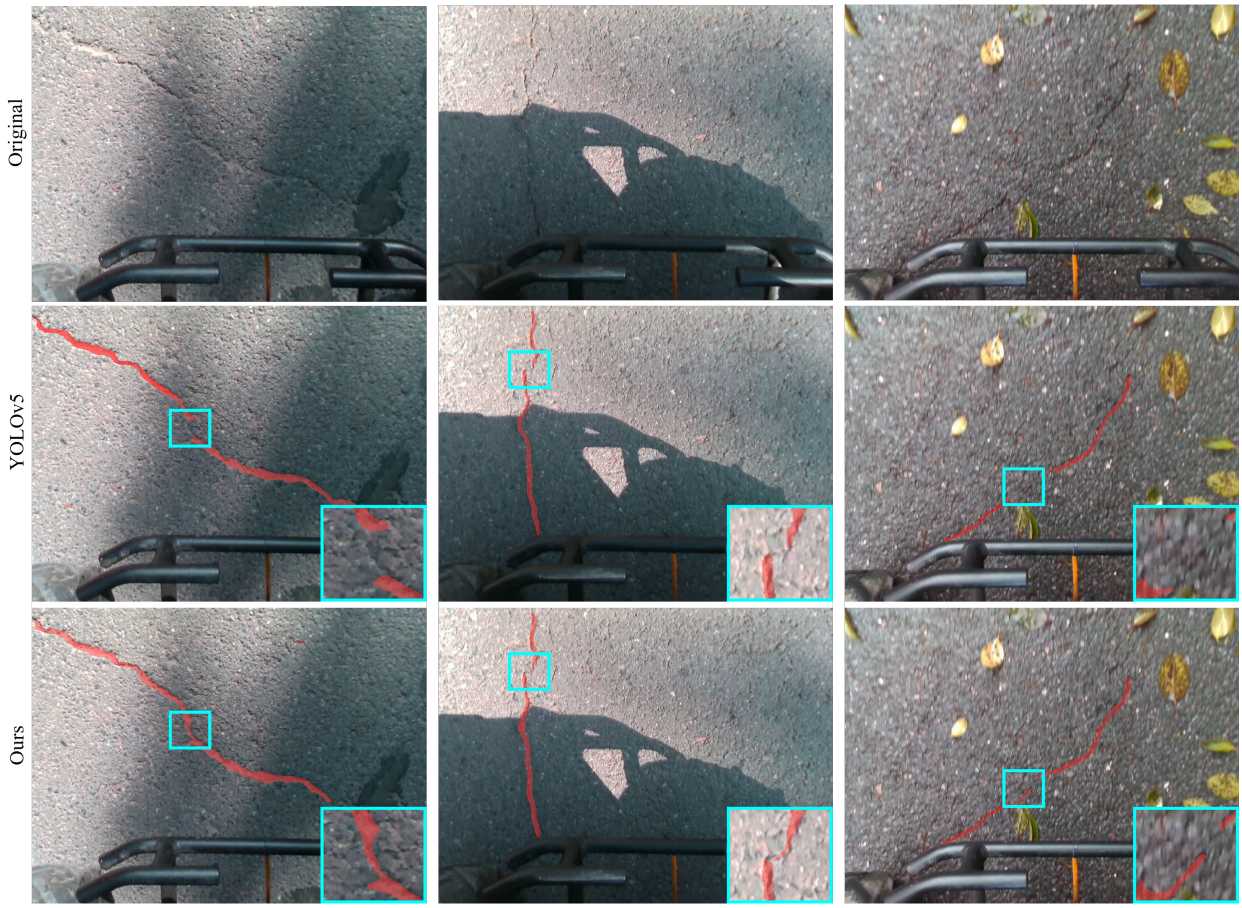

Figure 4. Visualization of segmentation results using YOLOv5 and our proposed method of our created S2T-Crack dataset.

According to the convergence curve, the Neural-PID algorithm needs 60 iterations to find the local optimal solution and 90 iterations to get rid of the local optimal solution. Compared with the 150 iterations required by the CEAFC method, this is a huge reduction. Therefore, Neural-PID can eliminate the local optimal solution and improve the robustness of crack tracking control.

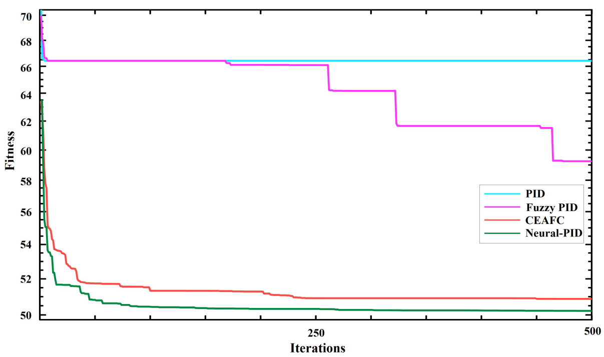

Figure 5. The comparison results of algorithm optimization.

4.3.2 Analysis of Tracking Error

Table 2. The comparison results of crack tracking error.

Crack ID	Segmentation Model	Control Method			
		PID	Fuzzy PID	CEAFC	Neural-PID
#1	n	9.71	5.81	4.68	4.47
	s	9.57	5.73	4.54	4.12
	m	9.93	6.07	4.73	4.59
	l	10.24	6.44	4.91	4.75
	x	11.86	6.76	5.16	5.01
#2	n	13.12	6.21	5.03	4.94
	s	12.86	6.19	4.85	4.63
	m	13.38	6.58	5.17	4.86
	l	13.89	6.91	5.43	5.21
	x	14.31	7.35	5.79	5.57
#3	n	15.08	7.73	6.25	5.94
	s	14.59	7.51	5.86	5.67
	m	15.36	8.09	6.57	6.29
	l	16.18	8.67	6.93	6.76
	x	16.85	7.28	7.06	6.81

Experiments are performed on real roads to verify the performance of road crack tracking, as shown in Table 2. This average absolute error is used as a performance evaluation index. The unmanned wheeled robot uses the proposed method to compare the results of road crack tracking error with other control methods during the tracking process. Crack #1 is a straight pavement crack. In the case of crack #1, our algorithm achieves the smallest average crack tracking absolute error in the pre-trained model ’s’, with a measured value of 4.12 mm. Crack #2 is a curved pavement crack. For the case of crack #2, our algorithm achieves the smallest average absolute error in the pre-trained model ’s’, with a measured value of 4.63 mm. Crack #3 is a continuous turning pavement crack. Our algorithm achieves the minimum mean absolute error in the pre-trained model ’s’, and the measured value is 5.67 mm.

5 Conclusions

This article addresses two critical issues in road crack tracking: insufficient feature extraction and low tracking efficiency. To overcome these challenges, the research primarily focuses on enhancing the pavement crack feature extraction from crack video images using our transformer-based crack segmentation method. By combining SA and CA, and leveraging FT model, the performance of binary masks in segmentation instances is significantly improved, enabling fine-grained segmentation of pavement cracks. Through the proposed Neural-PID, our method is deployed on NVIDIA AGX Xavier to enable real-time tracking of actual pavement cracks on a vision-guided robot. In future research, the utilization of road crack depth images will be considered, along with the exploration of alternative control methods to enhance the accuracy and robustness of the tracking control algorithm. The developed vision-guided robot can be integrated with repair mechanisms to accomplish road crack repairs.

Acknowledgements

The study presented in the article was partially supported by the National Key Research and Development Program of China (No.2021YFB2601000), National Natural Science Foundation of China (No.52078049, No.52378431), Natural Science Foundation of Shaanxi Province (2022JM-193), Fundamental Research Funds for the Central Universities, CHD (No.300102210302, No.300102210118), the 111 Project of Sustainable Transportation for Urban Agglomeration in Western China (No.B20035).

References

[1] Jingwei Liu, Xu Yang, Stephen Lau, Xin Wang, Sang Luo, Vincent Cheng-Siong Lee, and Ling Ding. Automated pavement crack detection and segmentation based on two-step convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, 35(11):1291–1305, 2020. doi:10.1111/mice.12622.

[2] Jingwei Liu, Xu Yang, Xin Wang, and Jian Wei Yam. A laboratory prototype of automatic pavement crack sealing based on a modified 3D printer. *International Journal of Pavement Engineering*, 23(9):2969–2980, 2022. doi:10.1080/10298436.2021.1875225.

[3] Jinchao Guan, Xu Yang, Ling Ding, Xiaoyun Cheng, Vincent C.S. Lee, and Can Jin. Automated pixel-level pavement distress detection based on stereo vision and deep learning. *Automation in Construction*, 129: 103788, 2021. doi:10.1016/j.autcon.2021.103788.

[4] Jinchao Guan, Xu Yang, Pengfei Liu, Markus Oeser, Han Hong, Yi Li, and Shi Dong. Multi-scale asphalt pavement deformation detection and measurement based on machine learning of full field-of-view digital surface data. *Transportation Research Part C: Emerging Technologies*, 152:104177, 2023. doi:10.1016/j.trc.2023.104177.

[5] Zhihao Pan, Jinchao Guan, Xu Yang, Kang Fan, Jeremy C.H. Ong, Ningqun Guo, and Xin Wang. One-stage 3D profile-based pavement crack detection and quantification. *Automation in Construction*, 153:104946, 2023. doi:10.1016/j.autcon.2023.104946.

[6] Jianqi Zhang, Xu Yang, Wei Wang, Jinchao Guan, Ling Ding, and Vincent C. S. Lee. Automated guided vehicles and autonomous mobile robots for recognition and tracking in civil engineering. *Automation in Construction*, 146:104699, 2023. doi:10.1016/j.autcon.2022.104699.

[7] Chengjia Han, Tao Ma, Ju Huyan, Xiaoming Huang, and Yanning Zhang. CrackW-Net: A Novel Pavement Crack Image Segmentation Convolutional Neural Network. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):22135–22144, 2022. doi:10.1109/TITS.2021.3095507.

[8] Guijie Zhu, Jiacheng Liu, Zhun Fan, Duan Yuan, Peili Ma, Meihua Wang, Weihua Sheng, and Kelvin C. P. Wang. A lightweight encoder–decoder network for automatic pavement crack detection. *Computer-Aided Civil and Infrastructure Engineering*, pages 1–23, 2023. doi:10.1111/mice.13103.

[9] Frank K.A. Awuah and Alvaro Garcia-Hernández. Machine-filling of cracks in asphalt concrete. *Automation in Construction*, 141:104463, 2022. doi:10.1016/j.autcon.2022.104463.

[10] Jianqi Zhang, Xu Yang, Wei Wang, Jinchao Guan, Wenbo Liu, Hainian Wang, Ling Ding, and Vincent C. S. Lee. Cross-entropy-based adaptive fuzzy control for visual tracking of road cracks with unmanned mobile robot. *Computer-Aided Civil and Infrastructure Engineering*, pages 1–20, 2023. doi:10.1111/mice.13108.

[11] Firdes Çelik and Markus König. A sigmoid-optimized encoder–decoder network for crack segmentation with copy-edit-paste transfer learning. *Computer-Aided Civil and Infrastructure Engineering*, 37(14):1875–1890, 2022. doi:10.1111/mice.12844.

[12] Xinzi Sun, Yuanchang Xie, Liming Jiang, Yu Cao, and Benyuan Liu. DMA-Net: DeepLab With Multi-Scale Attention for Pavement Crack Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):18392–18403, 2022. doi:10.1109/TITS.2022.3158670.

[13] Bo Chen, Hua Zhang, Guijin Wang, Jianwen Huo, Yonglong Li, and Linjing Li. Automatic concrete infrastructure crack semantic segmentation using deep learning. *Automation in Construction*, 152:104950, 2023. doi:10.1016/j.autcon.2023.104950.

[14] Honghu Chu and Pang-jo Chun. Fine-grained crack segmentation for high-resolution images via a multiscale cascaded network. *Computer-Aided Civil and Infrastructure Engineering*, pages 1–20, 2023. doi:10.1111/mice.13111.

[15] Dongho Kang, Sukhpreet S. Benipal, Dharshan L. Gopal, and Young-Jin Cha. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. *Automation in Construction*, 118:103291, 2020. doi:10.1016/j.autcon.2020.103291.

[16] Zhihan Zhang and Di Bai. Optimization of Improved PID Control Strategy Based on Genetic Algorithm. *Journal of Physics: Conference Series*, 2417(1):012025, 2022. doi:10.1088/1742-6596/2417/1/012025.

[17] YeFei Kang, ZhiBin Li, and Tao Wang. Application of PID Control and Improved Ant Colony Algorithm in Path Planning of Substation Inspection Robot. *Mathematical Problems in Engineering*, 2022:1–10, 2022. doi:10.1155/2022/9453219.

[18] Ultralytics. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. `https://github.com/ultralytics/yolov5.com`, 2022. URL `https://doi.org/10.5281/zenodo.7347926`. Accessed: 7th May, 2023.

[19] Yuchuan Du, Shan Zhong, Hongyuan Fang, Niannian Wang, Chenglong Liu, Difei Wu, Yan Sun, and Mang Xiang. Modeling automatic pavement crack object detection and pixel-level segmentation. *Automation in Construction*, 150:104840, 2023. doi:10.1016/j.autcon.2023.104840.

[20] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, Zhaohui Yang, Yiman Zhang, and Dacheng Tao. A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):87–110, 2023. doi:10.1109/TPAMI.2022.3152247.

[21] Yuqing Wang, Zhaoliang Xu, Xinlong Wang, Chunhua Shen, Baoshan Cheng, Hao Shen, and Huaxia Xia. End-to-End Video Instance Segmentation with Transformers. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8737–8746, Virtual, Online, United states, June 2021. IEEE Computer Society. doi:10.1109/CVPR46437.2021.00863.

[22] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip H.S. Torr, and Li Zhang. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6877–6886, Virtual, Online, United states, June 2021. IEEE Computer Society. doi:10.1109/CVPR46437.2021.00681.

[23] Wael Farag. Complex Trajectory Tracking Using PID Control for Autonomous Driving. *International Journal of Intelligent Transportation Systems Research*, 18(2):356–366, 2020. doi:10.1007/s13177-019-00204-2.

[24] Salman Bari, Syeda Shabih Zehra Hamdani, Hamza Ullah Khan, Mutte Ur Rehman, and Haroon Khan. Artificial neural network based self-tuned PID controller for flight control of quadcopter. In *2019 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–5, Lahore, Pakistan, February 2019. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/CEET1.2019.8711864.