MECHANICAL ENGINEERING SCIENCE

ISSN:2661-4448(online)
2661-443X(print)
Volume 5 No.2 2023

VISER
www.viserdata.com

Mechanical Engineering Science

VISER

Publisher: Viser Technology Pte. Ltd.

ISSN: 2661-4448(online)

2661-443X(print)

Frequency: Semi-annual

Add.: 21 Woodlands Close, #08-18

Primz Bizhub SINGAPORE (737854)

https://www.viserdata.com/

Editors:

Yajun LIAN Yanli LIU

John WILSON Nike Yough

Mart CHEN Qiuyue SU

Debra HAMILTON Xin DI

Jennifer M DOHY Xiuli LI

Edward Adam Davis

Designer: Anson CHEE

Mechanical Engineering Science

Volume 5 No.2 (2023)

CONTENTS

VISER

Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12716

Research Article **Open Access**

A Review of Research on Person Re-identification in Surveillance Video

Yunzuo ZHANG[*], Weiqi LIAN

School of Information science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043, China

[*]***Corresponding Author:*** *Yunzuo ZHANG, E-mail: zhangyunzuo888@sina.com*

Abstract

Person re-identification has emerged as a hotspot for computer vision research due to the growing demands of social public safety requirements and the quick development of intelligent surveillance networks. Person re-identification (Re-ID) in video surveillance system can track and identify suspicious people, track and statistically analyze persons. The purpose of person re-identification is to recognize the same person in different cameras. Deep learning-based person re-identification research has produced numerous remarkable outcomes as a result of deep learning's growing popularity. The purpose of this paper is to help researchers better understand where person re-identification research is at the moment and where it is headed. Firstly, this paper arranges the widely used datasets and assessment criteria in person re-identification and reviews the pertinent research on deep learning-based person re-identification techniques conducted in the last several years. Then, the commonly used method techniques are also discussed from four aspects: appearance features, metric learning, local features, and adversarial learning. Finally, future research directions in the field of person re-identification are outlooked.

Keywords: *Person re-identification; Deep learning; Metric learning; Local features; Adversarial learning*

1 Research Background and Significance of Person Re-identification

In recent years, with the acceleration of urbanization and the growth of population size, urban traffic, public safety and other issues have become incrSeasingly prominent. Traditional manual patrols and monitoring means have been difficult to meet the demand, and the public's demand for public safety has been increasing. A significant number of surveillance cameras have been placed in public areas including streets and neighborhoods as a result of the ongoing advancements in surveillance video intelligent analysis technologies. And these cameras cover a wide geographic range, generating a large amount of surveillance video data [1]. Person re-identification (Re-ID) [2] technology research has become an inevitable trend in order to study the behavioral traits and activity patterns of individuals in these surveillance recordings in a timely and efficient manner.

Person re-identification can realize automatic tracking and identification of persons, give full play to the role of surveillance video data, and improve the efficiency of the city in maintaining public safety and traffic order. The purpose of person re-identification is to find the image of a person with a given identity in a large-scale image library, and to determine whether the persons in different viewpoints, different cameras, and different video clips are the same person [3]. This technique is commonly used in applications like crowd behavior analysis, multi-camera target tracking, and target detection.

The key to person re-identification is to learn the discriminative features of individuals to differentiate between images of the same person and those of different individuals. However, in reality, people can appear in several cameras in various locations, and it is difficult to learn person discriminative features due to variations in different camera viewpoints and resolutions, person poses, and ambient lighting. Traditional methods for person re-identification primarily rely on manually extracting discriminative features or learning better similarity measures. However, these methods are prone to errors and are time-consuming, which significantly impacts the accuracy and real-time performance of the person re-identification task.

Automated person re-identification methods have significantly advanced the development of intelligent surveillance [4-5]. With automated person re-identification methods, surveillance systems can more accurately recognize the same person appearing in different cameras, reduce the rate of misrecognition, and improve the overall accuracy of surveillance systems [6]. This helps to detect abnormal behaviors or suspicious people in time and

Received on October 25, 2023; Accepted on November 29, 2023

improve the security of the surveillance system. In addition, automated person re-identification methods can improve the processing speed and real-time performance of the surveillance system, enabling the surveillance system to respond more quickly to unexpected events or security threats [7]. The automated person re-identification method can provide richer data and feature information for the surveillance system, supporting the system to perform deep learning and data analysis, so as to realize intelligent decision-making and prediction [8]. By analyzing data such as person activity patterns and paths, the surveillance system can better understand and predict potential risks and improve the intelligence of the surveillance system.

With the development of deep neural network (DNN) based techniques [9], person re-identification techniques are able to extract significant person features more efficiently. Deep neural networks can learn a high-level abstract representation of the data, which allows it to extract more robust pedestrian features under different lighting, angle and occlusion conditions. With an end-to-end training approach, DNNs can automatically learn feature representations that are best suited for person re-identification tasks [10]. Compared to traditional feature extraction methods, DNN can better perform nonlinear modeling, thus better capturing complex features and structural information in pedestrian images [11]. In addition, with a pre-trained deep neural network model, generalized features learned on large-scale image data can be migrated to the person re-identification task. This type of migration learning can accelerate the training process of the model and improve the performance of the model on small sample data [12].

2 Current Status of Domestic and Overseas Research

The traditional method of person re-identification is to manually extract a single fixed person feature, which is not only error-prone but also consumes a lot of time. The emergence of deep learning has greatly promoted the research and development in the field of person re-identification. Person re-identification technology based on deep learning has been applied in many fields.

Research on feature extraction has progressively moved from conventional manual feature extraction techniques to deep learning-based techniques [13]. In order to be able to extract more robust and detailed features, Sun et al. [14] introduced local feature learning and Tay et al. [15] used an attention mechanism to concentrate on the important details of body parts. By merging the global and local features of individuals, Zhang et al. [16] improved the final feature representation. Because Generative Adversarial Networks (GANs) are good at creating images and learning features, generative adversarial learning has been applied extensively in person Re-ID tasks [17-19]. Several loss functions have been created by some researchers to maximize the ability of network models to learn discriminative features [20-22].

When compared to still photos, video or image sequences provide additional information about human attributes. Therefore, examining these more comprehensive attribute signals could be a more effective approach to address the person re-identification challenge. Important identifying information lost in a single image due to occlusion, changes in field of view, and other causes can be compensated for by the temporal correlation between frames of a person sequence or trajectory. In view of these results, several researchers have begun to efficiently merge more information in video sequences by utilizing complementing spatial and temporal cues [23-25]. Furthermore, by modeling graph relations on person photos, several researchers have employed graph convolutional network-based techniques to learn more robust and discriminative features [26-28].

Three-dimensional shape information is an important cue for understanding person pose and shape. With the exception of the true 3D human anatomy and the spatial interaction between people and interfering objects, the majority of person Re-ID techniques now in use learn person feature representations from photos. The robustness of person Re-ID models can be improved by utilizing information about the 3D shape of the person [29-30].

3. Datasets and Evaluation Metrics for Person Re-ID

This section contains datasets that are frequently used in person Re-ID techniques that are based on deep learning. Furthermore, a brief description of popular evaluation measures for person Re-ID is provided.

3.1. Commonly used datasets

DukeMTMC-VideoReID [31] consists of approximately 4,832 videos from 1,812 identities, using Duke as its acronym. There are 702 identities in the dataset for testing and training, and 408 identities for distraction. Manual annotations are made to its bounding boxes.

The first extensive dataset for video-based re-identification was introduced in 2016 and is called MARS [32]. It comprises over 20,000 video sequences of person trajectories that were obtained from six different cameras, as well as 1,261 identity sequences. Sixty-five of these twelve hundred IDs were utilized for training, while the remaining six hundred and thirty-six were used for testing.

CUHK01 [33] comprises 3,884 manually cropped photos and 971 IDs, with each individual having at least two images captured using two separate cameras. In the CUHK01 dataset, the images from one of the cameras have more different viewpoints and pose variations, while the images from the other camera mainly consist of front and back views.

With 1,816 subjects and 7,264 manually cropped photos, CUHK02 [34] includes ten camera perspectives total—five pairs of views. At least two pictures in two separate discontinuous camera viewpoints have been taken of each participant. This dataset has more person identities and camera views than CUHK01, and more person picture attributes are available for retrieval.

CUHK03 [13] is a large Re-ID dataset. Ten cameras provide data for CUHK03, which detects person bounding boxes using hand labeling and a deformable part model (DPM) detector. It has 13,164 photos total, each of varied sizes, and 1,360 distinct people. CUHK03 captures more photographs and has more cameras in order to record people from a variety of angles.

Table 1 Commonly used datasets for person re-identification

Datasets	Data type	ID	Boxes/ Tracks	Cameras	Labeled
DukeMTMC-VideoReID	Video	1,812	4,832	8	DPM
MARS	Video	1,261	20,715	6	DPM+GMMCP
CUHK01	Image	971	3,884	2	Handcrafted
CUHK02	Image	1,816	7,264	10	Handcrafted
CUHK03	Image	1,360	13,164	10	DPM+ Handcrafted
Market-1501	Image	1,501	32,217	6	DPM+ Handcrafted
MSMT17	Image	4,101	126,441	15	Faster RCNN

Five high-resolution cameras and one low-resolution camera were used to collect Market-1501 [35]. Market-1501 automatically detects the human bounding box using the person detector DPM. It comprises 32,668 photos total, each measuring 128 by 64, and it features 1,501 distinct people. Market-1501 features more annotated photos and an interference factor, making the graphics more realistic than CUHK03.

MSMT17 [36] from which 15 cameras provided the images. For automatic detection, it makes use of the person detector Faster R-CNN labeled frames. This is one of the larger person and annotated image datasets for the current person Re-ID endeavor, with 4,101 distinct person details and 126,441 photos. The MSMT17 dataset has a larger range of perspectives, considerable illumination differences, and a greater coverage of scenes.

Table 1 shows the details of the above datasets. Many methods are challenging to apply in real-world applications due to uncertainty caused by occlusion, lighting changes, camera view switching, position changes, and related clothing. In person re-identification research, it is imperative to investigate large-scale person Re-ID datasets covering more real-world events.

3.2. Commonly used evaluation metrics

The commonly used evaluation indexes for evaluating the algorithm model in the person re-identification technology are CMC [37] curve and mAP [38].

CMC curve is a very important and commonly used evaluation index in pattern recognition, which can comprehensively reflect the performance of person Re-ID model. CMC uses the top-k approach, which involves sorting the results based on how similar the query target and the target image are to one other. Top-k is the likelihood that, upon sorting, the first k photos returned have the right query result. To arrive at the final result, the hit probability of each query image that was previously collected is added up and divided by the total number of query images.

The accuracy and recall of each category are combined by mAP. In a target detection task, objects are usually categorized according to categories, and for each category, its AP [39] value can be calculated. The mAP value is then calculated by averaging these AP values. In general, the detection method performs better when the mAP is higher. For each category, the detection results are sorted from high to low confidence, and then the number of True Positive (TP), False Positive (FP) and False Negative (FN), i.e., the confusion matrix, is determined based on the threshold value. The confusion matrix for binary classification is shown in Table 2. Accuracy and recall are calculated based on the number of TP, FP and FN, and AP values are calculated for different recall rates. Lastly, the mAP value is calculated by averaging the AP values across all categories. The mAP evaluation metric is commonly used in person re-identification to comprehensively assess algorithm performance while accounting for differences between multiple categories.

Table 2 Binary confusion matrix

	Prediction=1	Prediction=0
Reference=1	TP	FN
Reference=0	FP	TN

4 Deep Learning Based Person Re-identification Methods

In this section, deep learning-based person re-identification methods are classified into four categories. As shown in Figure 1, these categories include appearance-based features, metric-based learning, localized feature-based, and adversarial learning-based methods. In addition, the above four methods are introduced and their respective advantages and disadvantages are discussed.

4.1 Methods based on appearance feature

This type of method mainly performs re-identification by extracting appearance features of persons, such as color, texture and shape. These are shallow visual characteristics, and convolutional neural networks (CNN), local binary patterns (LBP), and the histogram of oriented gradients (HOG) are frequently used feature extraction techniques. Liao et al. [40] presented a person re-identification technique with

LOMO+XQDA, which improved the RANK-1 accuracy on four datasets.

In addition to this, person re-identification can also be performed based on information such as backpacks, glasses, hair, and so on. And this information belongs to the mid-level visual features, i.e., semantic attributes. The semantic attributes of the same person change very little under different camera shots. Therefore, these semantic attributes can be combined with the attribute weights and shallow features to finally describe the person image.

Feature extraction techniques will directly affect the accuracy of identification, and establishing high-level visual features is a challenge in appearance feature extraction. Matsukawa et al. [41] proposed GOG, which divides an image into horizontal strips and local blocks. Each strip is modeled with a Gaussian distribution and establishes high-level visual features by extracting image regions with predefined block or stripe shapes in color or texture histograms.

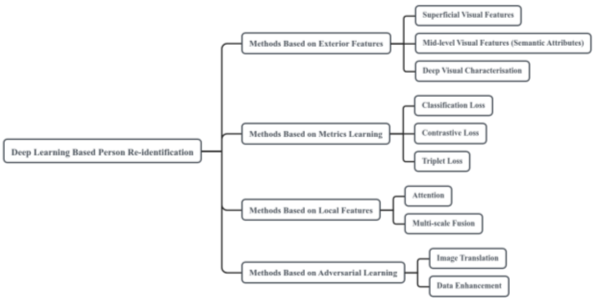


Figure 1 Classification of Deep Learning Based Person Re-identification Methods

4.2 Methods based on metric learning

This type of method mainly measures the similarity between different persons by learning a suitable distance metric. Person re-identification is a common use of deep metric learning (DML).

Among these, marginal distance, cosine similarity, and Euclidean distance are the most often utilized metric learning techniques. While the loss function of the model is typically calculated to constrain the learning of discriminative features in deep metric learning [42]. Commonly used loss functions include classification loss [43-45], contrastive loss [46-47] and triplet loss [48-49]. Because classification loss has the benefits of straightforward training and mining hard samples, it is commonly utilized in person re-identification algorithms based on metric learning. Models that are not sufficiently capable of generalization are not learned by using ID information alone. Thus, in order to limit the model's training, classification loss typically has to be paired with other losses. The contrastive loss is a good representation of sample pair matching.In models for feature extraction, the contrastive loss is frequently employed in conjunction with the classification loss to train the network. Deep learning based Re-ID techniques are frequently used in conjunction with triplet loss and classification loss to enhance model performance. The goal of these techniques is to reduce the intra-class distance and maximize the inter-class distance of the samples.

In addition to this, some studies propose new ways of computing the loss function to improve the modeling approach [50]. Du et al. [51] proposed an adaptive heterogeneous centroid loss, which reduces the distance between samples corresponding to the same identity while increasing the centroid distance for different samples. By eliminating the need for expensive manual feature design, the deep metric learning approach enables the model to automatically learn for feature recognition.

4.3 Methods based on local features

Person re-identification techniques can be classified as global or local based on the features that are taken from person photos for categorization.

Global feature-based methods usually extract only one feature of the persons [52], which makes the feature information of the persons we can get is limited. In order to obtain more detailed, nuanced, and discriminative features, local feature-based methods have gained widespread attention. Learning people's local features and maintaining their alignment are the goals of local feature-based learning approaches. These techniques concentrate on an individual's head, upper body, or lower body, among other local aspects. Person re-identification models can perform better by extracting and matching local features.

Paying attention to every element could be very expensive. An attention method can be used to solve this issue by concentrating just on the local feature regions that have a significant influence. In order to better cluster people's temporally dispersed features and remove contaminated features, Wang et al. [53] introduced a temporal attention mechanism and proposed a new temporal attention module that adaptively evaluates the importance of each extracted feature and combines all the valuable features together. While large-scale feature maps can more accurately depict the image's semantic content and contain the contour information of the image, small-scale feature maps contain more detailed information. The extracted multi-scale person features are combined using a multi-scale fusion technique to get richer person features. In order to overcome the challenges associated with cross-scale feature learning, Chen et al. [54] suggested a deep pyramid feature learning approach for learning multi-scale complimentary features. An all-scales network (OSNet) design was proposed by Zhou et al. [55] in order to learn features. The design encapsulates the synergistic combination of various sizes in addition to capturing aspects at distinct spatial scales.

4.4 Methods based on adversarial learning

The GAN generative adversarial network consists of a generator and a discriminator. The generator is responsible for generating pseudo-labels close to the real situation, while the discriminator is responsible for discriminating the pseudo-labels generated by the generator. The two fight against each other, constantly updating and iterating so that the results generated by the generator are closer and closer to the real value, and the discriminator is able to discriminate and distinguish the results that are closer to the real situation.

There is a domain gap between different datasets,i.e., there is a degradation of model performance when the same model is trained and tested on different datasets. In order to tackle this issue, Ge et al. [56] utilized an image-to-image translation network trained to translate images in two directions using the appropriate generators, utilizing the popular CycleGAN architecture. Additionally, by diversifying the training data, the model's capacity for generalization can be strengthened.An Adaptive Memory with Group Labeling (AdaMG) framework for Unsupervised Person Re-ID was proposed by Peng et al. [57]. It uses adaptive memory to build a multi-branch structure that resists noisy labels and takes use of data diversity. Dual et al. [58] proposed a framework which includes two phases of offline clustering, which refines the pseudo-labels, and online training, which optimizes the features. For person re-identification in UDA, Li et al. [59] suggested a triple adversarial learning and multi-perspective creative inference network that improves the discriminability and robustness of the learnt features while discriminating domain-invariant features.

5 Research Outlook

This paper provides an overview and summary of the literature on deep learning-based person re-identification technologies. Firstly, typical datasets and assessment indices for person re-identification technologies are presented, followed by a summary of common research directions in person re-identification both domestically and internationally. Secondly, four perspectives—appearance-based features, metric-based learning, localized features, and adversarial learning—are explored regarding the popular deep learning-based person re-identification techniques.

Some current person re-identification datasets suffer from category imbalance and inaccurate labeling, which may lead to degradation of model performance and generalization ability. Secondly, person re-identification techniques need to cope with the great variation of pedestrian appearance in different scenarios. The appearance of pedestrian images may vary greatly due to a variety of factors such as the angle of the surveillance camera, lighting conditions, pedestrian posture and occlusion. The robustness and generalization ability of current person re-identification techniques still need to be improved. In addition, the computational efficiency and storage space issues when dealing with large-scale data are also challenges for person re-identification techniques. In practical applications, surveillance systems often generate a large amount of pedestrian images and video data, which require efficient algorithms and sufficient storage space to support if they are to be comprehensively analyzed and processed. In real-world scenarios, there are often multiple pedestrian targets that need to be detected and recognized at the same time, which requires algorithms that can accurately locate and recognize multiple pedestrian targets in complex backgrounds.

In summary, although the existing research on deep learning-based person re-identification techniques has achieved very significant results. However, there are still many aspects that need attention and improvement in the future.

(1) Cross-modal person re-identification: Data in the real world may come from different types of cameras. For example, at night, ordinary cameras cannot clearly capture the features of persons, so infrared images from infrared cameras are needed to capture the features. When matching images from different modalities, reducing the differences in feature distribution in different modalities is a future task to be considered.

(2) 3D person re-identification: 3D shape information provides insight into richer person pose and shape information. The majority of current algorithms ignore the true 3D human anatomy and the spatial relationship between people and interferences in favor of just learning person features from photos. By obtaining 3D point cloud data of an image and processing it, data enhancement is performed by aligning 2D and 3D feature information.

(3) Domain adaptive person Re-ID: It becomes more challenging to effectively apply a Re-ID model that was trained on one dataset to another due to domain variations between datasets. While optimizing with pseudo-labels, a variety of unsupervised domain-adaptive techniques tend to transfer the learned knowledge from one domain to another, they are not without limits. In particular, these algorithms introduce a high number of noisy labels by one-time clustering, which hinders the retraining process and limits generalization, and they always generate a pseudo-label for each unlabeled sample, which makes it challenging to accurately describe an individual.

(4) Semi-supervised or unsupervised person re-identification: Usually the size of the dataset used by the model is very large and the overhead of annotating the data is very high. Therefore semi-supervised or unsupervised methods become a focus for future attention. They can extract features from unlabeled or partially labeled datasets. Therefore, in order to enhance the performance of semi-supervised or unsupervised algorithms in the future, more suitable clustering or label assignment procedures must be discovered.

References

[1] Wang Xiaogang. Intelligent multi-camera video surveillance: A review [J]. Pattern Recognition Letters, 2013,34:3-19.

[2] Ning Xin, Gong Ke, Li Weijun, et al.. Feature refinement and

filter network for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,31:3391-3402.
[3] Luo Hao, Jiang Wei, Fan Xing, et al.. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019,45:2032-2049.
[4] Bharath Ramesh, Shihao Zhang, Hong Yang,et al.. e-TLD: Event-based framework for dynamic object tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,31:3996-4006.
[5] Fan Heng, Ling Haibin, Siamese cascaded region proposal networks for real-time visual tracking[C]. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019.
[6] M. Saquib Sarfraz, Arne Schumann, Andreas Eberle,et al.. A pose-Sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
[7] Félix Remigereau;Djebril Mekhazni;Sajjad Abdoli, et al.. Knowledge distillation for multi-target domain adaptation in real-time person re-identification[C]. IEEE International Conference on Image Processing (ICIP), 2022.
[8] Wu Fei ,GaoYang ,Liu Jing ,et al.. Multi-stage feature interaction network for masked visible-thermal person re-identification[C]. 2023 35th Chinese Control and Decision Conference (CCDC), 2023.
[9] Pan Honghu ,Chen Yongyong ,He Zhenyu,et al..TCDesc: Learning topology consistent descriptors for image matching[J]. IEEE Trans. Circuits Syst. Video Technol, 2021,32:2845-2855.
[10] Yang Xi , Liu Liangchen, Wang Nannan , Gao Xinbo. A two-Stream dynamic pyramid representation model for video-based person re-identification[J]. IEEE Transactions on Image Processing, 2021,30:6266-6276.
[11] G. Han, M. Lin, Z. Li, H. Zhao and S. Kwong. Text-to-Image Person Re-identification Based on Multimodal Graph Convolutional Network[J]. IEEE Transactions on Multimedia, 2023,47:145-165.
[12] Zhang Xinyu,Cao Jiewei , Shen Chunhua ,et al.. Self-Training With Progressive Augmentation for Unsupervised Cross-Domain Person Re-Identification[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
[13] Wei Li,Rui Zhao,Tong Xiao,Xiaogang Wang. Deepreid: deep filter pairing neural network for person re-identification[C], IEEE International Conference on Computer Vision, 2014.
[14] Sun Tifan, Zheng Liang, Yang Yi,et al.. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]. European Conference on Computer Vision, 2018.
[15] C.P. Tay, S. Roy, K.H. Yap. Aanet: attribute attention network for person re-identifications[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[16] Zhang Hongwei ,Zhang Guoqing,Chen Yuhao,et al.. Global relation-Aware contrast learning for unsupervised person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:8599-8610.
[17] C. Wei, H. Wang, W. Shen,et al.. CO2: Consistent contrast for unsupervised visual representation learning[J]. 2020.
[18] Zhang Demao,Zhang Zhizhong, Ju Ying,et al.. Dual mutual learning for cross-modality person re-identification[J]. IEEE Trans. Circuits Syst Video Technol, 2022(47):550-560.
[19] E. Ristani, F. Solera, R. Zou, et al.. Performance measures and a data set for multi-target, multi-camera tracking[C]. Eur. Conf. Comput. Vis, 2016.
[20] Zheng Meng, S. Karanam, Wu Ziyan, et al.. Re-identification with consistent attentive siamese networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[21] Chai Tianrui, Chen Zhiyuan, Li Annan, et al.. Video person re-identification using attribute-enhanced features[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:7951-7966.
[22] Chen Weihua, Chen Xiaotang, Zhang Jiangguo,et al.. Beyond triplet loss: a deep quadruplet network for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[23] Yang Jinrui, Zheng Wei-Shi, Yang Qize, et al.. Spatial-temporal graph convolutional network for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.
[24] X. Liao, L. He, Z. Yang,et al.. Video-based person re-identification via 3d convolutional networks and non-local attention[C]. Asian Conference on Computer Vision, 2019.
[25] Yan Yichao, Qin Jie, Chen Jiaxin, et al.. Learning multi-granular hypergraphs for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.
[26] Bai Zechen, Wang Zhigang, Wang Jian, et al.. Unsupervised multi-source domain adaptation for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2021.
[27] Wu Yiming, O. Bourahla, Li Xi, et al.. Adaptive graph representation learning for video person re-identification[J]. IEEE Trans. Image Process, 2020,29:8821-8830.
[28] Yan Yichao, Zhang Qiang, Ni Bingbing,et al.. Learning context graph for person search[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[29] Huang Yewen, Lian Sicheng, Zhang Suian, et al..Three-dimension transmissible attention network for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,30:4540-4553.
[30] Wang Changshuo, Ning Xin, Li Weijun,et al.. 3D Person Re-identification Based on Global Semantic Guidance and Local Feature Aggregation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023(55):102-113.
[31] Wu Yu, Lin Yutian , Dong Xuanyi ,et al.. Exploit the unknown gradually:one-shot video-based person re-identification by stepwise learning[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[32] Zheng Liang, Bie Zhi, Sun Yifan,et al.. Mars: a video benchmark for large-scale person re-identification[C]. European Conference on Computer Vision, 2016.
[33] Li Wei, Wang Xiaogang. Locally aligned feature transforms across views[C]. IEEE Conference on Computer Vision and

Pattern Recognition, 2013.
[34] Li Wei, Zhao Rui, Wang Xiaogang. Human re-identification with transferred metric learning[C]. Asian conference on Computer Vision, 2012.
[35] Zheng Liang, Shen Liyue, Tian Lu, et al..Scalable person reidentification: a benchmark[C], IEEE International Conference on Computer Vision, 2015.
[36] Wei Longhui, Zhang Shiliang, Gao Wen, Tian Qi. Person transfer gan to bridge domain gap for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[37] Gray D, Tao H, Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]. European Conference on Computer Vision(ECCV), 2008.
[38] L. Zheng, L. Shen, L. Tian, et al.. Scalable person re-identification: a benchmark[C]. IEEE International Conference on Computer Vision (ICCV) 2015.
[39] Kunho Kim,Min-Jae Kim,Hyungtae Kim,et al..Person re-identification method using text description through CLIP[C]. 2023 International Conference on Electronics, Information, and Communication (ICEIC), 2023.
[40] Liao Shengcai, Hu Yang, Zhu Xiangyu,et al.. Person re-identification by local maximal occurrence representation and metric learning[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[41] MATSUKAWA T, OKABE T, SUZUKI E, et al.. Hierarchical gaussian descriptor for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[42] Ye Mang, Shen Jianbing, Lin Gaojie,et al..Deep learning for person reidentification: a survey and outlook[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2021(37):99-111.
[43] Zhong Zhun, Zheng Liang, Luo Zhiming, et al.. Invariance matters: exemplar memory for domain adaptive person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[44] Zhong Zhun, Liang Zheng, Zheng Zhedong,et al.. Camera style adaptation for person reidentification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[45] Huang Houjing, Li Dangwei, Zhang Zhang,et al.. Adversarially occluded samples for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[46] Wang Yicheng, Chen Zhenzhong, Wu Fen, et al.. Person re-identification with cascaded pairwise convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[47] M. Wang, B. Lai, and J. Huang, Camera-aware proxies for unsupervised person re-identification[C]. AAAI Conf. Artif. Intell. 2021.
[48] E. Ristani, C. Tomasi. Features for multi-target multi-camera tracking and re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[49] Song Chunfeng, Huang Yan, Ouyang Wanli,et al.. Mask-guided contrastive attention model for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[50] Zhang Peng, Xu Jingsong, Wu Qiang,et al..Top-Push constrained modality-adaptive dictionary learning for cross-modality person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020,30:4554-4566.
[51] Du Guodong, Zhang Liyan, Enhanced invariant feature joint learning via modality-invariant neighbor relations for cross-modality person re-Identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023(28):66-75.
[52] Wang Faqiang, W. Zuo, L. Lin,et al.. Joint learning of single-image and crossimage representations for person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[53] Wang Zhikang,He Lihuo,Tu Xiaoguang,et al.. Robust video-based person re-identification by hierarchical mining[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:8179-8191.
[54] Chen Yanbei, Zhu Xiatian, Gong Shaogang. Person re-identification by deep learning multi-scale representations[C]. IEEE International Conference on Computer Vision Workshops, 2017.
[55] Zhou Kaiyang,Yang Yongxin,Andrea Cavallaro.Learning generalisable omni-scale representations for person re-identification[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2021(26):66-71.
[56] Ge Yixiao, Zhu Feng, Chen Dapeng,et al.. Structured domain adaptation with online relation regularization for unsupervised person Re-ID[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024,35:258-271.
[57] Peng Jinjia, Jiang Guangqi, Wang Huibing. Adaptive memorization with group labels for unsupervised person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023,33:5802-5813.
[58] Dai Yongxing, Liu Jun, Bai Yan,et al.. Dual-refinement: Joint label and feature refinement for unsupervised domain adaptive person re-identification[J]. IEEE Trans. Image Process, 2021,30:7815-7829.
[59] Li Huafeng, Dong Neng, Yu Zhengtao, et al.. Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022,32:2814-2830.

Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12717

Research Article Open Access

Effects of Material Parameters on Stress Distribution in Casing-cement -formation (CCF) Multilayer Composite System

Chao ZHANG[1], Yuanbo XIA[2, 3], Bo ZHOU[1*], Xiuxing ZHU[1], Haijing WANG[1]

1. College of Pipeline and Civil Engineering, China University of Petroleum (East China), Qingdao, Shandong, 266580, China
2. School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, Shandong, 266580, China
3. CNPC Tianjin Bo-Xing Engineering Science &Technology Co., Ltd., Tianjin, 300451, China

**Corresponding Author:* *Bo ZHOU, E-mail: zhoubo@upc.edu.cn*

Abstract
This work focus on the stress distribution of the casing-cement -formation (CCF) multilayer composite system, which is a borehole system with multiple casings and cement sheathes. Most of the previous relevant studies are based on the traditional CCF system with the single casing and cement sheath, but these results are not adaptive to the CCF system multiple composite system. In this paper, the FEM numerical model of CCF multilayer composite system was constructed. Numerical simulations were calculated and compared with the system which consists of the single casing and cement sheath. Results show that the multilayer composite system possesses better performance. On this basis, the sensitivity analysis of main influence mechanical parameters such as in-situ stress, the elastic of cement sheathes and the elastic of formation are conducted. The cement sheath on the inside, namely cement sheath-1, is sensitive to its elastic modulus; meanwhile, the cement sheath on the outside, namely cement sheath-2, is not so sensitive to the elastic modulus of cement sheath-1. Cement sheath-1 and cement sheath-2 are all sensitive to the elastic modulus of cement sheath-2, and the mises stress of them has opposite trend to the elastic modulus of cement sheath-2. The proper values of elastic modulus of cement sheath-1 and cement sheath-2 are 5GPa and 5GPa to 30GPa, respectively. Under the in-situ stress ratio σh / σH = 0.7, the maximum mises stress of cement sheath-1 and cement sheath-2 increase as the increase of σh, and they are nearly equal when σh=15GPa. This research can be helpful for the design and analysis of CCF multilayer composite system.
Keywords: *In-situ stress; Stress distribution; Casing; Cement sheath; Formation; Multilayer*

1 Introduction

Casing-cement-formation (CCF) system is the borehole system commonly used for oil and gas exploitation operations in the petroleum industry [1-3]. This system is integral to maintaining borehole stability, ensuring the efficient production of hydrocarbons and providing a critical barrier against the uncontrolled flow of fluids between subsurface formations. However, when the system is subjected to the in-situ stress and internal pressure of drilling fluid, the mechanical integrity, especially the cement sheath, is often damaged. Then the formation fluid migrates from cracks and leading to the failure of annulus sealing [4-5]. Therefore, the research on the stress field of the CCF system is crucial to petroleum engineering.

The analytical method is one of the main methods to obtain the stress distribution of cement sheathes and casings. Yin et al. [6-7] formulated the analytical elastic solutions for a borehole system with the single of casing and cement sheath under the uniform and non-uniform in-situ stress, respectively. Besides the elastic analysis, Zhang et al. [8] derived the elastoplastic solutions of cement sheath under varied casing pressure. Considering the anisotropy of formation, Wang et al. [9-10] derived analytical solutions for cased borehole stress calculation in general anisotropic formations, which is closer to reality. In view of the far-field displacement boundary of the model of CCF is mostly not fixed, Yu et al. [11] proposed a modified model with a fixed far-field displacement boundary condition. In addition, considering the complexity of analytical solutions, Zhou et al. [12] proposed a semi-analytical method, it should be noticed that this method can not only simplify the calculation but also applicable for the borehole with multiple casings and cement sheathes.

In order to get the experimental data, many scholars conducted mechanical experiments about CCF system [13-15]. Bu et al. [16] designed a device for testing the interface radial bond strength, whose test results can be helpful to study the sealing integrity of cement sheath. Wu et al. [17] employed the digital image correlation (DIC) technique to examine the strain distribution and failure of cement sheath. Li et al. [18] studied the failures of ordinary cement sheath and expanding cement sheath through indoor experiments. Yan et al. [19] conducted a perforating

Received on November 3, 2023; Accepted on December 15, 2023

experiment in real size to investigate the damage mechanism of cement sheath. With the development of computer technology, many numerical methods have been favored by scholars which overcome the limit of the expensive experimental costs and the intricacy of analytical method. Fortunately, there are some available numerical methods to solve the practical problems, such as finite difference method (FDM) [20], finite element method (FEM) [21-22], boundary element method (BEM) [23], meshless methods [24] and so on.

The mentioned above just concentrated on the CCF system with the single casing and cement sheath. As the appearance of deep wells and ultra-deep wells [25-27], the operational environments and conditions becoming more complex, which is a greater challenge to the integrity of the CCF system. The CCF composite system which consists of multiple casings and cement sheathes is an efficient structure in deep wells and ultra-deep wells [28-29]. As for the existing researches for CCF multilayer composite systems, Zhou et al. [12] proposed the semi-analytical method which can conveniently obtain the stress distribution on every casing and cement sheath; based on the relevant theories of heat transfer and elastoplastic mechanics, Li et al. [33] formulated the thermal stress field of CCF multilayer composite systems; Huang et al. [34] conducted the application of hydra-jet fracturing technology in CCF with three layers casings, which provide the detailed references to simulate such well as with multi-layer casings.

However, most of the previous studies are based on the traditional CCF system with the single casing and cement sheath, which are not adaptive to the CCF multiple composite system. In this paper, the numerical model is constructed, which can be used for the comprehensive analyses of CCF multilayer composite system, and the stress fields of casings, cement sheathes and formation are calculated. A series of sensitivity analysis of main parameters are carried out, the results are valuable and helpful to design the CCF multilayer composite system. The rest of this paper is as follows: Firstly, the theoretical analysis of CCF multilayer composite system is elaborated in Section 2. Secondly, the basic theory of FEM correlation is stated in Section 3. Next, the FEM numerical model of CCF multilayer composite system is constructed and the comprehensive analyses of stress field are conducted in Section 4. Finally, this work is summarized, and some conclusions are concluded in Section 5.

2 Mechanical Analysis of CCF Multilayer Composite System

Figure 1 shows a casing-cement-formation (CCF) multilayer composite system, which subjected to the non-uniform in-situ stress. The CCF multilayer composite system can be seemed as a multilayer thick-walled cylinder, in which the inner pressure of the innermost casing, p_w, is induced by drilling fluid. The inner radius of the innermost casing is denoted by r_0, and the outer radius of other casings and cement sheathes are denoted by r_i (i = 1, 2, …, n) in sequence. From the innermost layer, the elastic modulus and Poisson's ratios of the casings and cement sheathes are denoted by E_i (i = 1, 2, …, n) and μ_i (i = 1, 2, …, n) in sequence, respectively. And the elastic modulus and Poisson's ratios of the formation are denoted by E_{n+1} and μ_{n+1}, respectively.

In this section, the model of the casing-cement-formation multilayer composite system is simplified to a plane strain problem, and the following assumptions are satisfied: (1) the casings, cement sheathes and formation are uniform and isotropic linear elastic materials; (2) the casings, cement sheathes and borehole are concentric; (3) the bonding surface between casings, cement sheathes and formation are completely consolidated.

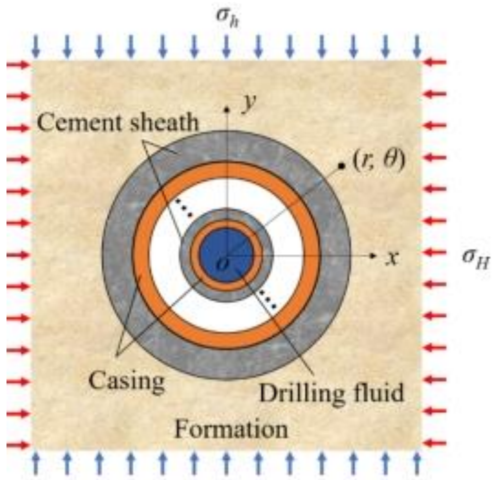


Figure 1 Diagram of CCF multilayer composite system, where σ_H is the maximum horizontal in-situ stress, and σ_h is the minimum horizontal in-situ stress

According to coordinate transformation in the elasticity theory [30], the in-situ stress shown in Figure 1 can be decomposed into the spherical stress tensor read as Eq. (1) and the deviatoric stress tensor read as Eq. (2).

After the decomposition of non-uniform in-situ stress shown in Figure 2, the initial problem can be transferred as the sum of two subproblems, one is the uniform load problem that the CCF multilayer composite system subjected to the inner pressure of drilling fluid and uniform stress shown in Eq. (2); the other is the system subjected to the deviatoric stress shown in Eq. (1). According to the superposition principle [7], the stress field of the CCF multilayer composite system, shown in Figure 1, should be the sum of the uniform stress field formulated by Eq. (3) and the deviatoric stress field formulated by Eq. (4), then the stress field of CCF multilayer composite system can be obtained.

$$s = \frac{1}{2}(\sigma_H - \sigma_h) \tag{1}$$

$$\sigma = \frac{1}{2}(\sigma_H + \sigma_h) \tag{2}$$

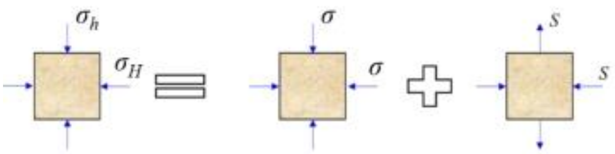


Figure 2 Decomposition of in-situ stress tensor, where (a) is the in-situ stress tensor, (b) is the spherical stress tensor, (c) is the deviatoric stress tensor

According to the reference [31], the solution of the subproblem of uniform load is as

$$a_i P_{i+1} - b_i P_i + c_i P_{i-1} = 0, (i = 1, \ldots, n-1) \tag{3}$$

where a_i, b_i and c_i are coefficients related to geometric and material parameters of casings, cement sheathes and formation; P_i is contact pressure of the contact surface between casings, cement sheathes and formation. When i=1, $P_{i-1}=P_w$; when i=n-1, $P_{i+1}=\sigma$. Based on the Eq. (3), an equation set with n-1 unknows about $P_1 \sim P_n$ can be obtained, thus the contact pressure of each contact surface can be solved. Then the stress distribution of casings and cement sheathes can be obtained by substituting the contact pressure into the Lamey's formula [30],

$$\left.\begin{aligned} \sigma_r &= -\frac{\frac{r_2^2}{r^2}-1}{\frac{r_2^2}{r_1^2}-1} q_1 - \frac{1-\frac{r_1^2}{r^2}}{1-\frac{r_1^2}{r_2^2}} q_2 \\ \sigma_\theta &= \frac{\frac{r_2^2}{r^2}+1}{\frac{r_2^2}{r_1^2}-1} q_1 - \frac{1+\frac{r_1^2}{r^2}}{1-\frac{r_1^2}{r_2^2}} q_2 \end{aligned}\right\} \tag{4}$$

where r_2 represents the outer radius; r_1 represents the inner radius; q_1 represents the inner pressure; q_2 represents the outer pressure.

For the subproblem of cosine load, the stress distribution of each contact interface is as

$$\left.\begin{aligned} (\sigma_r)_i &= -(2B_i + 4C_i r^{-2} + 6F_i r^{-4}) \cos 2\theta \\ (\sigma_\theta)_i &= (12A_i r^2 + 2B_i + 6F_i r^{-4}) \cos 2\theta \\ (\tau_{r\theta})_i &= (6A_i r^2 + 2B_i - 2C_i r^{-2} - 6F_i r^{-4}) \sin 2\theta \end{aligned}\right\} \tag{5}$$

where A_i, B_i, C_i, F_i are coefficients related to the geometric and material parameters of casings, cement sheathes and formation. Using displacement and stress continuity conditions and stress boundary conditions, the undetermined coefficients in Eq. (5) casings, cements and formation can be obtained.

3 FEM Basically Principle

The FEM is an efficient numerical method to find approximated solutions of the field variables in the problem domain. According to the FEM theory, the problem domain is firstly divided into several elements, as shown in Figure 3, and the connecting points between the elements are called nodes; then the field variables to be seek are set as the interpolation function of nodes; based on the variational principle, the problem can be transformed into a system of algebraic equations where the field variables of all nodes are unknown; finally, the boundary conditions are used to solve the system of algebraic, and the approximate solution of the field variables are obtained subsequently.

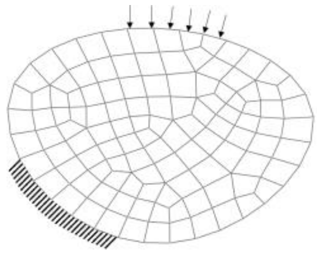


Figure 3 A FEM discrete structure of two-dimensional elasticity problem, which includes nodes and elements

For the two-dimension elasticity domain, the unknown field variable is displacement component. The displacement of a random point c in the domain shown in Figure 4 can be expressed as

$$u(\mathbf{x}) = \mathbf{p}^{\mathrm{T}}(\mathbf{x})\mathbf{b} \tag{6}$$

where P(x) is the vector polynomial basis functions, whose number of monomials can vary from 3 to 10, as listed in Tab. 1; b is the vector of undetermined coefficients related to polynomial basis functions.

$$\mathbf{b} = [b_1, \ b_2, \ \cdots, \ b_m]^{\mathrm{T}} \tag{6.1}$$

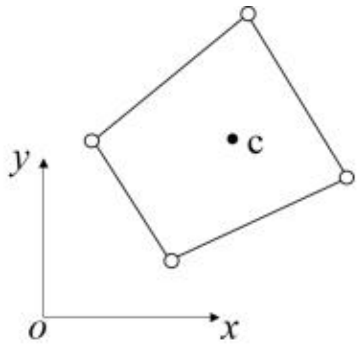


Figure 4 The random point c in the element

Table 1 The vectors of polynomial basis functions for two-dimensional domains, where m is the number of monomials [32]

m	p(x)
3	$[1, x, y]^T$
4	$[1, x, y, xy]^T$
5	$[1, x, y, x^2, y^2]^T$
6	$[1, x, y, x^2, xy, y^2]^T$
7	$[1, x, y, x^2, y^2, x^3, y^3]^T$
8	$[1, x, y, x^2, xy, y^2, x^3, y^3]^T$
9	$[1, x, y, x^2, y^2, x^3, x^2y, xy^2, y^3]^T$
10	$[1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3]^T$

Therefore, the displacement components of the point c shown in Figure 4 can be represented by that of the nodes on the element. So the displacement function can be expressed as

$$u(\mathbf{x}) = \mathbf{N}\mathbf{a}_e = \sum_{i=1}^{n} N_i u_i \tag{7}$$

where N is the vector of shape functions, N_i is the shape function related to node I; u_i is the nodal displacement.

Substituting the displacement function into the geometric function, we can obtain the element strain matrix. Further, the element stiffness matrix K_i is obtained according to the variational principle [32].

The relationship between nodal displacements and equivalent loads is expressed as

$$Ka=P \tag{8}$$

where K is the global stiffness matrix; P is the global vector of nodal equivalent loads; a is the global vector of nodal displacements, read as

$$\mathbf{a} = [u_1, v_1, u_2, v_2, \ldots, u_M, v_M]^T \tag{9}$$

M is the total number of nodes in the discrete structure shown in Figure 3. Solving the Eq. (8) can we obtain the value of total nodal displacement components. Then the nodal stress and strain components can be obtained by substituting the above values into the geometric function and physical function [32]. The global stiffness matrix is formulated as

$$\mathbf{K} = \sum_{i=1}^{N} K_i \tag{10}$$

where K_i is the elemental stiffness matrix, and N is the total number of elements of the discrete structure in Figure 4.

4 Numerical Analysis

In this section, a domestic oil well is taken to study the stress field of the CCF multilayer composite system, the structural diagram is shown in Figure 5(a). According to the basic principle of elastic mechanics, the mechanical problem of CCF multilayer composite system is the axisymmetric strain in the plane. Therefore, a quarter of the model is modeled to analyze the stress distribution. The FEM meshing model is shown in Figure 5(b). The inner walls between casings, cement sheathes and formation are all set to be welded. The minimum horizontal in-situ stress is applied to the upper boundary of the model; the maximum horizontal in-situ stress is applied to the right boundary; the normal displacement constraint is applied to the bottom and left boundary of the model, respectively.

The geometric and material parameters are listed in Table 2; the maximum of horizontal in-situ stress is 20 MPa; the minimum of horizontal in-situ stress is 13 MPa; the pressure of drilling fluid p_w is 30 MPa;

Table 2 Geometric and material parameters of casing-cement-formation multilayer composite system, where the r_{inner} represents the inner radius

Material	r_{inner}/mm	E/ GPa	μ
Casing-1	162.5	210	0.21
Cement sheath-1	178.3	8.5	0.3
Casing-2	229.1	210	0.21
Cement sheath-2	248.0	8.5	0.3
Formation	298.5	3.1	0.25

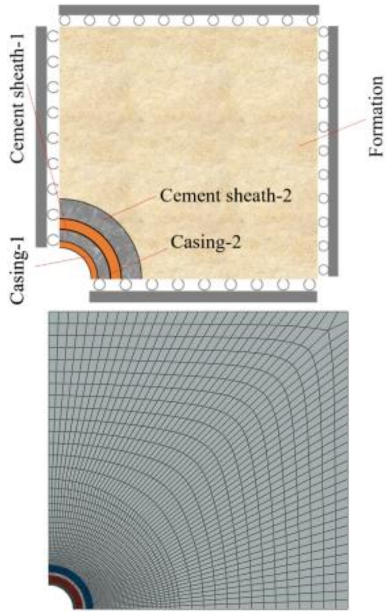


Figure 5 Physical model (a) and meshing (b) of casing-cement-formation multilayer composite system

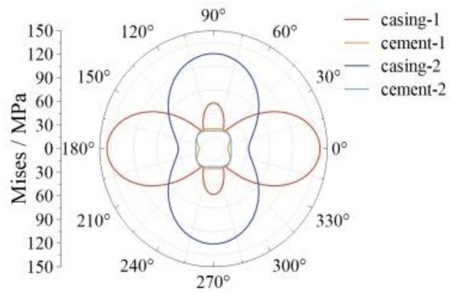


Figure 6 Mises stress distribution on the inner walls of multilayer composite

Figure 6 shows the mises distribution on the inner walls of casing-1, cement sheath-1, casing2 and cement sheath-2. It can be seen that the mises stress of casing-1 is larger than cement-1, and the mises stress of casing-2 is also larger than cement-2. This is because the elastic modulus of the casing is much larger than cement sheath. In addition, the mises stress distribution on the inner walls of casings and cements are non-uniform, which is caused by the deviatoric in-situ stress mentioned in Eq. 1. The above results are consistent in the mechanical analysis in section2, which verifies the numerical model of the CCF multilayer composite system.

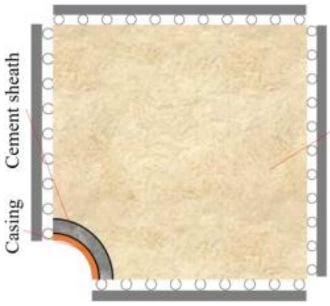


Figure 7 Physical model of casing-cement-formation single composite system

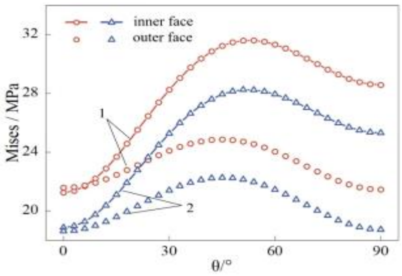


Figure 8 Mises stress distribution of inner and outer walls of cement sheathes, where 1 denote the cement sheath-1 in casing-cement-formation single composite system, and 2 denote the cement sheath in multilayer composite system

Based on the values listed in Tab. 2, we also calculated a single layer CCF system without casing 2 and cement sheath 2, as shown in Figure 7. The mises stress distribution of cement sheathes in multilayer composite system and single layer composite system were plotted in Figure 8. Figure 8 indicates that the mises stress of cement sheath in multilayer composite system is apparently significantly small than that in single composite system. Compared to the single composite system, the mises stress on the inner wall of cement sheath in multilayer composite system is decreased by 3.4MPa, and 7MPa decreased on the outer wall. Therefore, it is easy to find that the multilayer composite system can effectively enhance the stability of the wellbore.

In the rest of words, the sensitivity analysis of main influence mechanical parameters such as in-situ stress, the elastic of cement sheathes and the elastic of formation are conducted. Given the limited space, the cement sheathes shown in Figure 5(a), which is the most important and easily damaged in CCF system, is taken as the main study object.

4.1 Elastic modulus of formation

To investigate the effect of elastic modulus of formation, a range of sensitivity analysis conducted under the identical in-situ stress and modeling conditions. Figure 9 and Figure 10 shows the inner wall mises stress distribution of cement sheath-1 and cement sheath-2, respectively, for the values of elastic modulus of formation as 5GPa, 10GPa, 15GPa, 20GPa and 25GPa; Figure 10 presents the simulation results of the maximum mises stress versus the elastic modulus of formation.

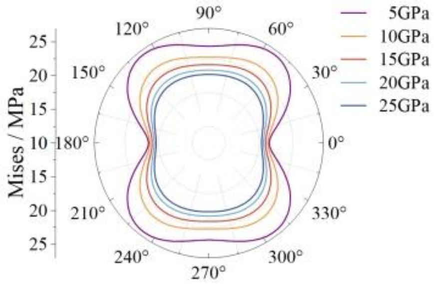


Figure 9 Mises stress distribution of cement sheath-1 inner wall under different elastic modulus of formation

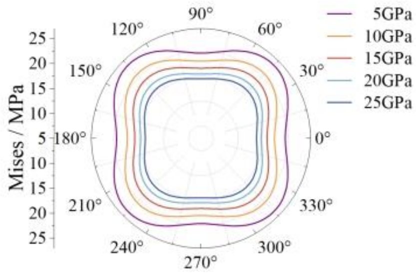


Figure 10 Mises stress distribution of cement sheath-2 inner wall under different elastic modulus of formation

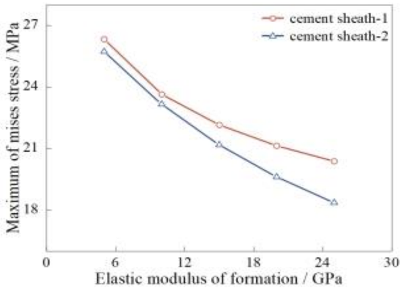


Figure 11 The variation of the maximum Mises stress on cement sheathes with different formation elastic modulus

As revealed from the results, with the elastic modulus of formation varying from 5GPa to 25GPa, the mises stress in the cement sheath-1 and 2 decreases, which indicates that the multilayer composite system is endurable in the formation of high elastic modulus. The maximum mises stress of cement sheath-1 is greater than that of cement sheath-2 under the identical elastic modulus of formation. The slope of the maximum mises stress curve of cement sheath-2 is larger than that of cement sheath-1, which demonstrates that the cement sheath-2 is more sensitivity to the differences of the elastic modulus of formation. Based on the above, it is advised that the CCF multilayer composite system is more suitable for stratum with high elastic modulus.

4.2 In-situ stress

In order to demonstrate the effects of the in-situ stress on the stress distribution of CCF multilayer composite system, several numerical simulations under different in-situ stress are conducted. Under the constant in-situ stress ratio $\sigma_h / \sigma_H = 0.7$, calculations are developed with several value sets of maximum horizontal and vertical in-situ stress listed in Tab. 3. Figure 12 and Figure 13 show the inner wall mises stress distribution of cement sheath-1 and cement sheath-2, respectively; Figure 14 illustra;5tes the simulation results of the maximum mises stress versus the in-situ stress.

According to the results form Figure 12 and 13, with the σ_h is altered form 3MPa to 15MPa, the mises stress in the cement sheath-1 and 2 increases. In Figure 14, the slope of the maximum mises stress curve of cement sheath-2 is larger than that of cement sheath-1, which indicates that the cement sheath-2 is more sensitivity to the differences of the in-situ stress under the constant in-situ stress ratio σ_h / σ_H = 0.7. Additionally, when σ_h varying from 3MPa to 15MPa, the maximum mises stress of cement sheath-1 is greater than that of cement sheath-2, and the difference between these tends to close as the σ_h increases. Based on the above, it is advised that more attention should be paid to the design of cement sheath-2 strength when the value of in-situ stress is too high.

Table 3 In-situ stress of formation, where σ_H is the maximum horizontal in-situ stress, σ_h is the maximum vertical in-situ stress

Number	σ_H / MPa	σ_h / MPa
1	4.3	3
2	8.6	6
3	12.9	9
4	17.1	12
5	21.4	15

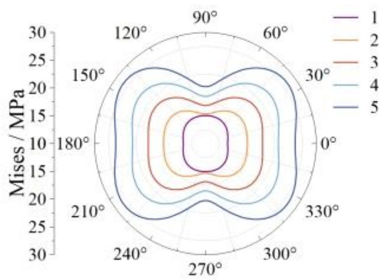


Figure 12 Mises stress distribution of cement sheath-1 inner wall under different in-situ stress, where 1, 2, 3,4 and 5 represent the in-situ set number in Table 3

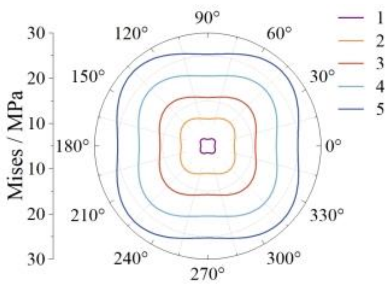


Figure 13 Mises stress distribution of cement sheath-2 inner wall under different in-situ stress, where 1, 2, 3,4 and 5 represent the in-situ set number in Table 3

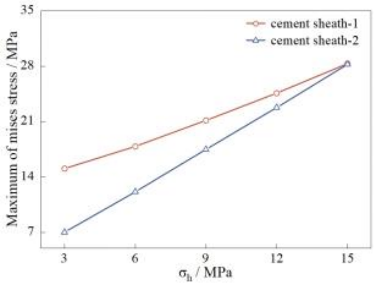


Figure 14 The variation of the maximum Mises stress on cement sheathes under different formation in-situ stress listed in Table 3

4.3 Elastic modulus of cement sheath

For the CCF multilayer composite system, the mechanical parameter of material is one of keys to the mechanical performance. Actually, the mechanical properties of casing are generally constants in engineering, and the main difference between various cement sheath is elastic modulus. Therefore, the effect of elastic modulus of cement sheath on the multilayer composite system is mainly discussed in this section. A series of numerical simulations with different elastic modulus of cement sheath-1 and cement sheath-2 are conducted, respectively.

4.3.1 Cement sheath-1

Figure 15 and Figure 16 show the inner wall mises stress distribution of cement sheath-1 and cement sheath-2 when the elastic modulus of cement formation-1 is 5GPa, 25 GPa, 45 GPa and 65 GPa, respectively. Figure 17 exhibits the maximum mises stress versus the elastic modulus of cement sheath-1.

Figure 15 shows that the inner wall mises stress of cement sheath-1 decreases with the increase of elastic modulus of cement sheath-1. According to the Figure 16, when the elastic modulus of cement sheath-1 increases, the values of mises stress decreases in the well angle is 45 ° to 135 °, and increases in the well angle is 135 ° to 225 °. Meanwhile, the maximum value of cement sheath-2 stays around 27Mpa. Figure 17 presents the simulation results of the maximum mises stress versus the elastic modulus of cement sheath-1, it is noticeably observed that the maximum mises stress of cement sheath-1 is much more sensitive than that of cement sheath-2.

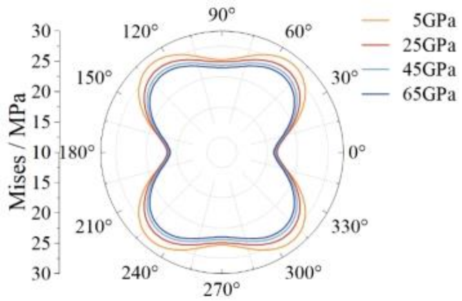


Figure 15 Mises stress distribution of cement sheath-1 inner wall under different elastic modulus of cement sheath-1

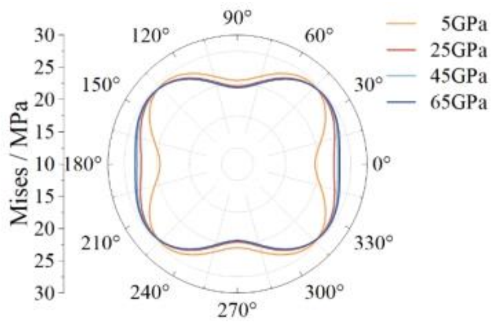


Figure 16 Mises stress distribution of cement sheath-2 inner wall under different elastic modulus of cement sheath-1

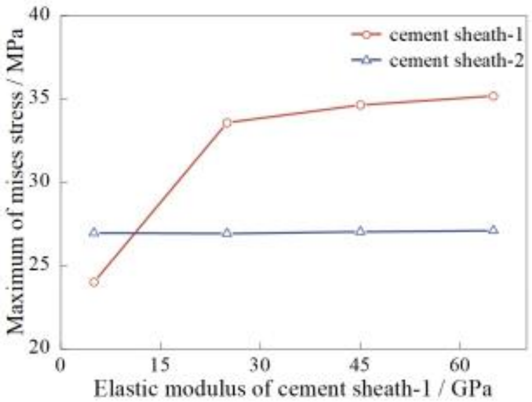


Figure 17 The variation of the maximum mises stress on cement sheathes under different elastic modulus of cement sheath-1

4.3.2 Cement sheath-2

Figure 18 and Figure 19 shows the inner wall mises stress distribution of cement sheath-1 and cement sheath-2 when the elastic modulus of cement formation-2 is 5GPa, 25 GPa, 45 GPa and 65 GPa, respectively. Figure 20 exhibits the maximum mises stress versus the elastic modulus of cement sheath-2.

As revealed form the results, the maximum mises stress of cement sheath-1 decreases with the increase of the elastic modulus of cement sheath-2. On the contrary, the maximum mises stress of cement sheath-2 increases with the increase of the elastic modulus of cement sheath-2.

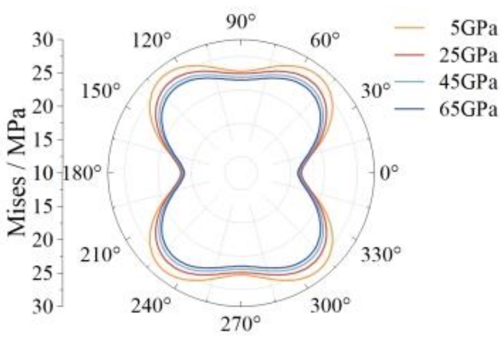


Figure 18 Mises stress distribution of cement sheath-1 inner wall under different elastic modulus of cement sheath-2

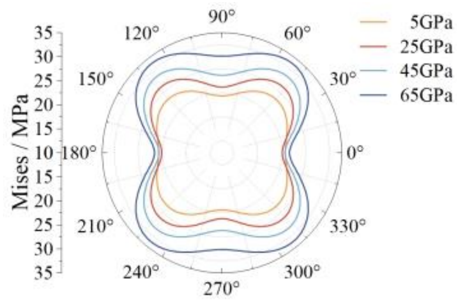


Figure 19 Mises stress distribution of cement sheath-2 inner wall under different elastic modulus of cement sheath-2

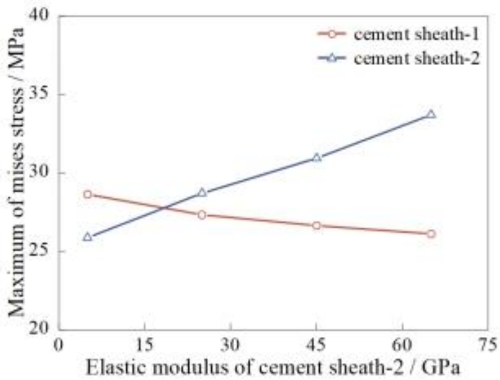


Figure 20 The variation of the maximum Mises stress on cement sheathes under different elastic modulus of cement sheath-2

Based on the comprehensive analysis of the elastic modulus of two cement sheathes, some suggestions are supposed that the elastic modulus of cement sheath-1 should be set around 5GPa and the elastic modulus of cement sheath-2 should be set around 5GPa to 30GPa. This setting could effectively improve the stability of the CCF multilayer composite system.

5. Conclusion

The theoretical analysis of casing-cement-formation multilayer composite system is elaborated. The FEM model of CCF multilayer composite system is constructed and a series of numerical simulations are conducted. Results indicate that the CCF multilayer composite system possesses better performance than the CCF system with single casing and cement sheath. Meanwhile, the sensitivity analysis which is the main influence of mechanical parameters of cement sheathes and formation is conducted.

In the inner wall of cement sheath-1 and cement sheath-2, the mises stress decreases as the elastic modulus of formation varying from 5GPa to 25GPa; and the mises stress increases as the in-situ stress increases, and the cement sheath-2 is more sensitivity to the in-situ stress under the constant in-situ stress ratio σ_h / σ_H = 0.7. Furthermore, the elastic modulus of cement sheath is essential for the stress distribution, in this paper, the proper values of elastic modulus of cement sheath-1 and cement sheath-2 are 5GPa and 5GPa to 30GPa, respectively. This research can be helpful to design the structure of CCF multilayer composite system.

Conflict of Interest: We are committed to that there is no conflict of interest regarding the publication of this paper.

Acknowledgments: The authors of this paper acknowledge the supports from the Independent Innovation Research Program of China University of Petroleum (East China) (Grant No. 27RA2215005) and the National Key Research and Development Program of China (Grant No. 2017YFC0307604).

This is the Appendix: This article does not cover the details that require an appendix.

References

[1] De Simone Marcelo, Pereira Fernanda L.G. and Deane Mesquita Roehl. Analytical methodology for wellbore integrity assessment considering casing-cement-formation interaction [J]. International Journal of Rock Mechanics and Mining Sciences, 2016, 94: 112-122.

[2] Wei Liu, Baohua Yu, and Jingen Deng. Analytical method for evaluating stress field in casing-cement-formation system of oil/gas wells [J]. Applied Mathematics and Mechanics, 2017, 38(9): 1273-1294.

[3] Yang Heng, Bu Yuhuan, Guo Shenglai, et al.. Effects of in-situ stress and elastic parameters of cement sheath in salt rock formation of underground gas storage on seal integrity of cement sheath [J]. Engineering Failure Analysis, 2021, 123: 105258.

[4] Guo Xinyang, Bu Yuhuan, Li Juan, et al.. Modes and prevention of cement sheath failures under complex downhole conditions [J]. Nature Gas Industry, 2013, 33(11): 86-91.

[5] Qu Zhan, Wang Xiaozeng and Dou Yihua. Analytic and Numerical Solutions of Load and Stress of Casing and Cement in Cementing Section [J]. Applied Mechanics and Materials, 2012, 268-270(268-270): 721-724.

[6] Yin Youquan, Chen Chaowei, Li Pingen. Theoretical solutions of stress distribution in casing-cement and stratum system [J]. Chinese Journal of Theoretical and Applied Mechanics, 2006, 28(6): 835-842.

[7] Yin Youquan, Cai Yongen, Chen Chaowei, et al.. Theoretical solution of casing loading in non-uniform ground stress field [J]. Acta Petrolei Sinica, 2006, 27(4): 133-138.

[8] Zhang Hong, Shen Ruichen, Li Jingcui, et al.. Elastoplastic Analysis of Cement Sheath of Injection Production Wellbore Under Varied Casing Pressure [J]. Arabian Journal for Science and Engineering, 2018, 43(11): 6523-6534.

[9] Wang Hongwei, Fang Xinding, Zhang Shuangxi, et al.. An analytical approach for cased borehole stress calculation in general anisotropic formations [J]. International Journal for Numerical and Analytical Methods in Geomechanics, 2021, 45(1): 3-27.

[10] Wang Hongwei, Miao Miao, Zhou Tao et al.. An analytical approach to the stress distribution around a cased borehole in an anisotropic formation [J]. Applied Mathematical Modelling, 2022, 107: 316-331.

[11] Yu Guimin, Xu Jie, Liu Wei, et al.. A Modified Model of Cement Sheath Stress Distribution with a Fixed Far-Field Displacement Boundary Condition [J]. Chemistry and Technology of Fuels and Oils, 2023, 59(2): 362-374.

[12] Zhou Bo, Zhang Chao, Zhang Xudong, et al.. A semi-analytical method for the stress distribution around a borehole with multiple casings and cement sheaths [J]. Geoenergy Science and Engineering, 2023, 229: 212091.

[13] Kuanhai Deng, Yue Yuan, Yi Hao, et al.. Experimental study on the integrity of casing-cement sheath in shale gas wells under pressure and temperature cycle loading [J]. Journal of Petroleum Science and Engineering, 2020, 195: 107548.

[14] Dou Haoyu, Dong Xuelin, Duan Zhiyin, et al.. Cement Integrity Loss due to Interfacial Debonding and Radial Cracking during COsub2/sub Injection [J]. Energies 2020, 13(17): 4589-4589.

[15] Maryam Tabatabaei, Arash Dahi Taleghani and Nasim Alem. Measurement of mixed mode interfacial strengths with cementitious materials [J]. Engineering Fracture Mechanics, 2020, 223: 106739.

[16] Bu Yuhuan, Tian Leiju, Guo Bingliang, et al.. Experiment and simulation on the integrity of cement ring interface in deep water shallow formation [J]. Journal of Petroleum Science and Engineering, 2020, 190: 107127-107127.

[17] Wu Yuxing, Harshkumar Patel, Saeed Salehi, et al.. Experimental and finite element modelling evaluation of cement integrity under diametric compression [J]. Journal of Petroleum Science and Engineering, 2020, 188: 106844-106844.

[18] Li Juan, Guo Xinyang, Bu Yuhuan, et al.. Investigation on preventing mechanical failure of cement sheath in fractured wells by expanding cement [J]. Oil Drilling & Production Technology, 2014, 36(4): 43-46.

[19] Yan yan, Guan Zhichuan, Wang qing, et al.. Experiment on damage of cement sheath induced by perforation in oil and gas wells [J]. Journal of China University of Petroleum. 2022, 46(3): 81-88.

[20] Kubacka Ewelina and Ostrowski Piotr. Heat conduction issue in biperiodic composite using Finite Difference Method [J]. Composite Structures, 2021, 261: 113310.

[21] Ma xiao and Wei Gaofeng. Numerical prediction of effective electro-elastic properties of three-dimensional braided piezoelectric ceramic composites [J]. Composite Structures, 2017, 180: 420-428.

[22] Zhou Bo, Kang Zetian, Wang Zhiyong, et al.. Finite element method on shape memory alloy structure and its applications. Chinese Journal of Mechanical Engineering, 2019, 32(84): 1-11.

[23] Sun Fenglong, Gong yanpeng and Dong Chunying. A novel fast direct solver for 3D elastic inclusion problems with the isogeometric boundary element method [J]. Journal of Computational and Applied Mathematics, 2020, 377: 112904.

[24] Ma Xiao, Zhou Bo and Xue Shifeng. A meshless Hermite weighted least-square method for piezoelectric structures[J]. Applied Mathematics and Computation, 2021, 400:1-15.

[25] Deng Hu and Jia Lichun. Key technologies for drilling deep

and ultra-deep wells in the Sichuan Basin: Current status, challenges and prospects [J]. Natural Gas Industry, 2022, 42(12): 82-94.

[26] Chang Deyu, Li Gensheng, Shen Zhonghou, et al.. The stress field of bottom hole in deep and ultra-deep wells [J]. Acta Petrolei Sinica, 2011, 32(4): 697-703.

[27] Wu Yi, Zhou Jianliang, Yang Jin, et al.. A Study on the Integrity Evaluation of Cement Sheaths for Deep Wells in Deep Water [J]. Energies, 2022, 15(16),5814-5814.

[28] Ding Shidong, Liu Kui, Liu Xiaogang, et al.. The Effect of Pre-Applied Annulus Back Pressure Cementing on Radial Stress of Interfaces in Double Layer Casing Systems [J]. Petroleum Drilling Techniques, 2022, 50(1): 30-37.

[29] Zhao Xinbo, Han Shengchao, Yang Xiujuan, et al.. Mechanical characteristics analysis of casing-cement sheath-formation multilayer composite system with thermo-structural coupling effects, 2017, 48(3): 837-843.

[30] Zhou Bo, Xue Shifeng and Zhu xiuxing. Solid Mechanics: Theory and MATLAB Solution [M]. Press of China University of Petroleum, Beijing China, pp. 49–70.

[31] Wang Yanbin, Gao Deli, Fang Jun. Mechanical characteristics analysis of casing-cement ring-formation multilayer composite system [J]. Chinese Journal of Applied Mechanics, 2014, 21(3): 387-392.

[32] Zhou Bo, Zhang Chao and Zhao Fei. A Finite Element-Meshless Hybrid Method (FEMLHM) of Elasticity Problem and Its Applications [J]. Mechanics of Solids, 2023, 58(3): 852-871.

[33] Li Yong, Ji Hongfei, Xing Pengju, et al.. Theoretical solutions of temperature field and thermal stress field in wellbore of a gas well [J]. Acta Petrolei Sinica, 2021, 42(1): 84-94.

[34] Huang Zhongwei, Li Gensheng, Wang Yongzhang, et al.. Hydra-jet fracturing applied in a well with three-layer casings [J]. Oil Drilling & Production Technology, 2012, 34(5): 122-124.

Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12718

Research Article Open Access

Simulation Study on Optical Properties of GaN-based Blue LED

Huilong BIAN[1],Taiping HAN[2*],Xiaoming GU[3]

1. Inner Mongolia Shenying Intelligent Technology Co., Ltd, Tongliao, Neimenggu, 028000, China
2. School of Materials Science and EngineeringXiamen University of Technology, Xiamen, Fujian, 361024, China
3. Tongliao Science and Technology Museum, Tongliao, Neimenggu, 028000, China

[*]***Corresponding Author:*** *Taiping HAN, E-mail: hantaiping21@163.com*

Abstract

The optical properties of GaN-based blue light-emitting diodes (LEDs) are extremely important to study as these LEDs are utilized in a great many industries due to their excellent qualities, including high brightness, high energy efficiency, low energy consumption, and rapid reaction time. In this paper, Silvaco TCAD simulation software is used to do two-dimensional modeling and simulation of a GaN-based blue single quantum well vertical structure LED, with an emphasis on varied forward voltages, In components in InGaN, and quantum well thickness. The volt-ampere characteristic curve is compared and evaluated, as well as the energy band structure, carrier concentration, radiation recombination efficiency, electroluminescence spectrum, and internal current density distribution. The results show that when the forward voltage is 3.5V and the thickness of the quantum well is constant, the luminescence spectrum will show a red shift with the increase of the In content in the quantum well, and the luminescence spectrum will also show a red shift when the thickness of the quantum well is increased. However, when the quantum well thickness and In component are kept constant, the luminescence spectrum appears a red shift with increasing forward voltage.

Keywords: *GaN-based; Blue light; LED; Spectrum; Silvaco TCAD*

1 Introduction

Light-emitting diode (LED) is a kind of semiconductor optoelectronic device that can convert electrical energy into light energy[1]. LED as a new energy-saving light source has many advantages: First, it has high efficiency and a long life; its power consumption is less than 10 times that of an incandescent lamp, and its service life can reach 100000 hours. Second, small size, fast response and rich colors (ultraviolet to infrared). Third, green environmental protection, free of mercury, lead and other heavy metal pollution. According to the relevant parameters, the forbidden band width of InN is 0.7 eV and that of GaN is 3.4 eV [2]. Different forbidden band thicknesses can be obtained by controlling different InN and GaN alloy components, and the forbidden band thicknesses range from 0.7eV to 3.4eV, and their wavelengths range from 365 nm to 1770 nm. By combining a blue LED light source with an already existing red LED light source and a green LED light source, the trichromatic idea may be used to generate a more realistic and practical white light[3-5]. InGaN-series materials used to make blue LEDs are widely employed in a variety of industries. Consequently, there has been a lot of interest in the research into the characteristics of blue GaN-based LEDs. The quantum wells formed in the energy band by InGaN and GaN materials, their thickness, and the amount of In components present will determine their luminous properties. Furthermore, the forward voltage, the amount of Al in the p-type AlGaN layer, the kind of substrate material, and the doping concentration all have a substantial impact on its optoelectronic properties[6]. In this study, a two-dimensional simulation model of a blue single quantum well LED based on GaN is created using the SilvacoTCAD simulation software's Atlas module, and the effects of forward voltage,In component in InGaN and quantum well thickness on a GaN-based blue single quantum well LED are explained and examined.

2 Simulation and Analysis

2.1 GaN-based LED structure

The typical structure of a GaN-based single quantum well LED is a diode composed of p-type semiconductors and n-type semiconductors, which are composed of direct energy gap semiconductors. In p-type semiconductors, holes are the most abundant carriers, and in N-type semiconductors, free electrons are the most abundant carriers[7-8]. The GaN layer is usually used as the electron barrier layer and the InGaN layer as the electron potential well layer. In the actual production process, sapphire is

Received on November 3, 2023; Accepted on December 10, 2023

commonly employed as the substrate material for GaN-based blue LEDs due to its high temperature resistance, high strength, good stability, and outstanding device quality. An InGaN/GaN single quantum well vertical structure LED on a sapphire substrate is used in this investigation. To simplify the model, a 300nm n-type GaN layer with a doping concentration of 1×1018 cm-3 is formed on a 600nm sapphire substrate, followed by a 3nm thick InGaN layer, as shown in figure 1. A 100 nm thick p-type AlGaN layer with a doping concentration of 1×1019 cm-3 is then created as an electron barrier layer (EBL), confining free electrons in the quantum well area and preventing free electrons from exiting the active region. After the inclusion of suitable Al components, EBL may limit the diffusion of Mg-doped in the p-type GaN layer to the quantum well region, enhancing the luminous efficiency and optical decay characteristics of LED. Finally, a 500 nm thick p-type GaN layer was produced with a doping concentration of 1×1019 cm-3. P-electrodes are made of the metal Ni, whereas N-electrodes are made of the metal Ti. Figure 2 depicts the quantum well band diagram. The energy band in the quantum well is tilted when the input voltage is applied, holes and free electrons accumulate in the quantum well region, and the accumulated holes and free electrons transition in the barrier layer, resulting in a recombination phenomenon that causes LEDs to emit light [9-13].

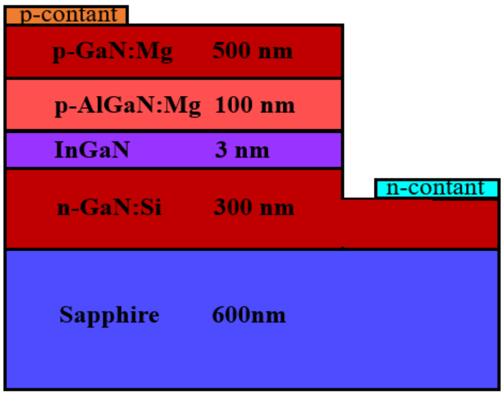


Figure 1 Schematic diagram of the basic structure of a GaN-based single quantum well LED

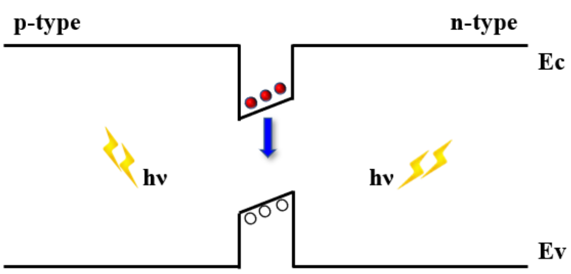


Figure 2 Energy band diagram of the quantum well

2.2 The influence of forward bias voltage

LED volt-ampere characteristics are also known as current-voltage characteristics. When the forward voltage applied to the LED is less than the turn-on voltage, it is difficult to overcome the potential barrier caused by the PN junction's built-in electric field, which hinders the diffusion movement of most carriers and produces high resistance and low current, which are insufficient to make the LED glow. The potential barrier created by the PN junction's internal electric field is completely offset when the forward bias voltage is higher than the turn-on voltage. As a result, the current increases quickly, most carriers' diffusion motion is obviously accelerated, and the LED is turned on [14]. According to the simulation findings, the opening voltage of the model's LED is around 3.2V, which corresponds to the real opening voltage range, as shown in figure 3.

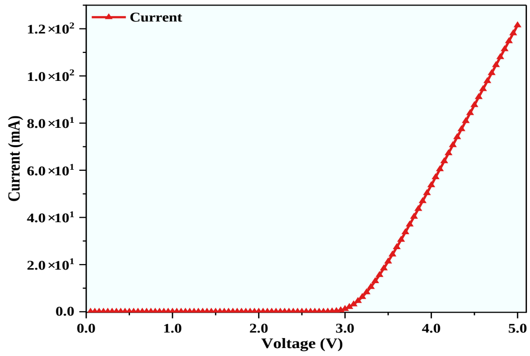


Figure 3 Volt-ampere characteristic curve of a GaN-based single quantum well LED

At forward voltages of 3.0V, 3.5V, 4.0V, 4.5V, and 5.0V, respectively, the electroluminescence spectra of an InGaN/GaN single quantum well vertical structure LED are shown in Figure 4. The LED light-emitting spectrum gains power density when the applied forward voltage is greater than the turn-on voltage, as shown in the figure. As the forward voltage rises, the peak wavelength gradually decreases, and the spectrum exhibits clear blue shift phenomena. Due to the carrier relaxation time in the conduction band (or valence band) being much shorter than the carrier lifetime, the carrier in the quantum well increases, shielding part of the built-in electric field, weakening the quantum-confined stark effect (QCSE), and the ground state in the well increases, so that the LED peak wavelength shifts to the short wave direction [15-16].

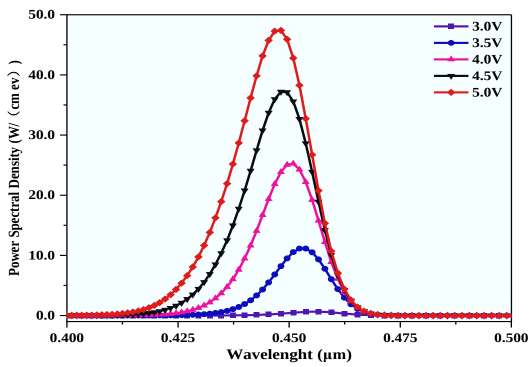


Figure 4 Electroluminescence spectra of GaN-based single quantum well LED structures change with forward voltage

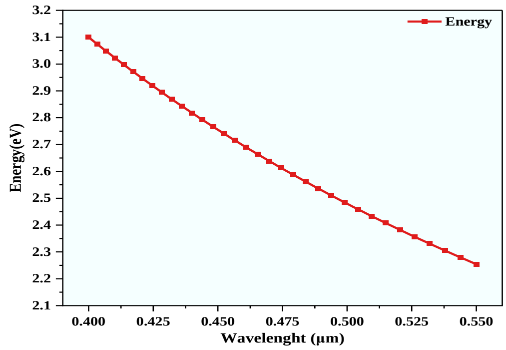


Figure 5 Relationship between the electroluminescence spectrum and energy of a GaN-based single quantum well LED structure

In figure 4, when the forward voltage increases to 3.5 V, the peak wavelength is 0.453 μm (453 nm). The simulation findings in figure 5 demonstrate that the energy is 2.73 eV, which is compatible with the theoretical estimates for the blue light peak wavelength range of 450 nm–490 nm and energy range of 2.53-2.76 eV. When the forward voltage reaches 5.0V, the current density progressively increases, the luminescence spectrum shifts toward the blue, the purple luminescence peak occurs at 0.44 μm, and the peak wavelength shortens. Figures 6(a)-(b) depict the flow direction of the device's internal current at 3.5 V and 5.0 V forward voltages, respectively. By enlarging the current cloud distribution map, it becomes clear that the current flows laterally in the lower portion of the n-type GaN layer and vertically downward in the p-type layer and well layer of the quantum well [17]. The current also flows vertically downward in the upper part of the p-type layer of the quantum well [18].

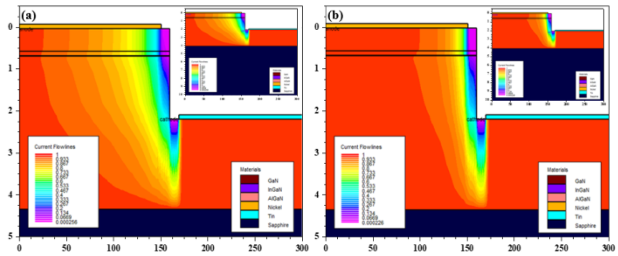


Figure 6 (a) Current density distribution cloud diagram of the LED model at forward voltage 3.5 V. (b) Current density distribution cloud diagram of the LED model at forward voltage 5.0 V

2.3 The influence of In component content

Figure 7 depicts the VI characteristic curve of a GaN-based LED with varied In components. The simulation results show that when the In content grows, the turn-on voltage fluctuates little, indicating that the In component has little impact on the turn-on voltage of the LED. In the case of keeping other parameters and the thickness of the quantum well constant, by varying the concentration of the In component, the effect of the In component on the performance of the GaN-based blue LED was investigated.

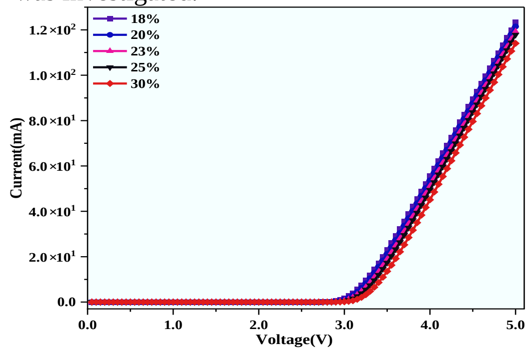


Figure 7 VI Characteristics of GaN-based LEDs with various In components

Figure 8 shows the energy band structures of different In components, which are affected by the piezoelectric effect and spontaneous polarization effect in LEDs, resulting in a polarized electric field. In addition, the lattice constant of InGaN/GaN does not match, and the increase of In content will increase the interface stress of the InGaN/GaN layer and enhance the polarization electric field, this causes the energy band in the quantum well structure to tilt [19]. The top of the valence band moves up, the bottom of the conduction band moves down, and the band gap width of the quantum well decreases. At the same time, the depth of the quantum well increases, thus increasing the relative barrier height, which increases the difficulty of carrier injection. Therefore, as the free electron and hole concentrations decline, so does radiation recombination efficiency, which in turn impacts LED luminous performance.

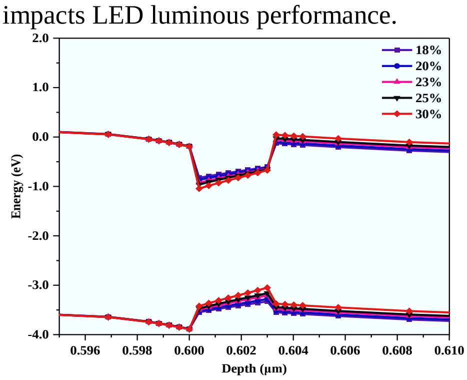


Figure 8 Energy band structure of quantum wells with different In components

Figure 9 shows the free electron and hole concentrations in GaN-based LED quantum wells with different In components. It can be seen from the simulation results that with the increase ofIn component, the free electron concentration near the n-type region decreases gradually and the hole concentration near the p-type region decreases gradually, which further confirms the results in figure 8, that is, as the free electron and hole concentrations in the quantum well decrease, the recombination efficiency is affected by the carrier concentration [20]. Figure 10 shows the radiation recombination efficiency in quantum wells with different In components. With the increase of In components, the radiation recombination efficiency decreases obviously, so the output efficiency is reduced, thus the quantum efficiency of LED is reduced[21].

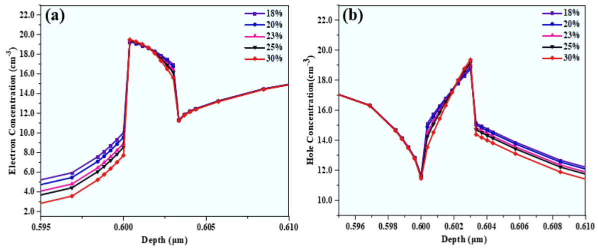


Figure 9 (a) Free electron concentration with different In components; (b) Hole concentration with different In components

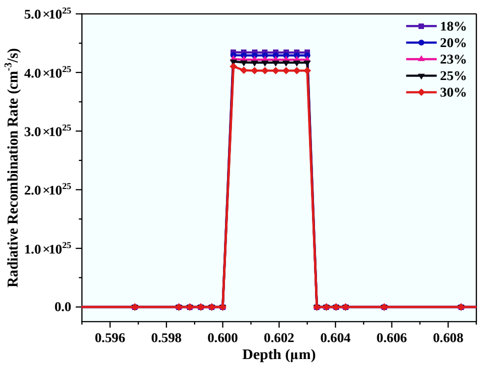


Figure 10 Radiation recombination efficiency with different In components

Figure 11 shows the relationship between optical power and current of LED with different In components. With the increase of In content, the corresponding optical power at the same current decreases. At the same time, with the increase of In content, the lattice matching of InGaN and GaN increases, the radiative recombination efficiency decreases, and the non-radiative recombination efficiency increases, which also leads to a decrease in optical output power.[22] As shown in figure 12, with the increase of the In component, the intensity of the power density decreases, the wavelength of the emitted photon increases gradually, and the peak wavelength shows a red shift to the long wave direction.

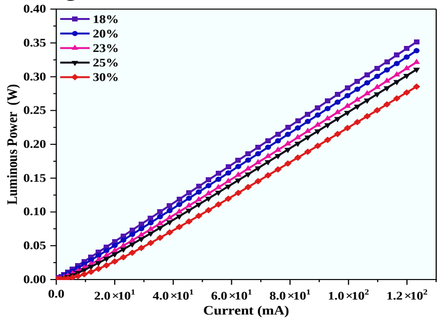


Figure 11 Relationship between optical power and current of LED with different In components

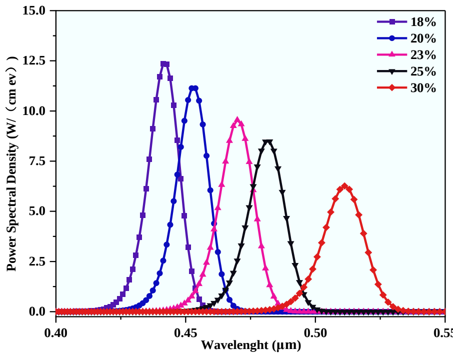


Figure 12 Changes of luminescence Spectra under different In components

2.4 The influence of the thickness of the quantum well

In the manufacturing process of LEDs, the thickness of the quantum well also affects their optoelectronic properties. The thickness range of a quantum well is generally 1－5 nm. The thickness of LED quantum dots well designed in this paper is 2nm, 3nm and 4nm. The thickness of the quantum well will affect the polarization electric field. Due to the existence of the polarization electric field, the energy band, radiation recombination efficiency and other parameters are affected. In the simulation, the working voltage of the LED is set to 3.5 V. Figure 13 shows the VI characteristic curve of GaN-based LEDs with different quantum well thicknesses. It can be seen from the figure that there is little difference in the turn-on voltage of the three structures, but from the illustration in figure 13, the LED turn-on voltage of the 2nm structure is the smallest and the LED turn-on voltage of the 4nm structure is the largest. The turn-on voltage increases with the thickness of the quantum well. Through the analysis, we believe that the main reason why the turn-on voltage of the device increases with the increase in the thickness of the quantum well is that with the increase of the thickness of the quantum well, the series resistance also increases, increasing the required opening voltage. When the thickness of the quantum well decreases, the greater the current obtained at the same voltage [23]. It can be seen from figure 14 that with the increase in the thickness of the quantum well, the tilt of the energy band increases.

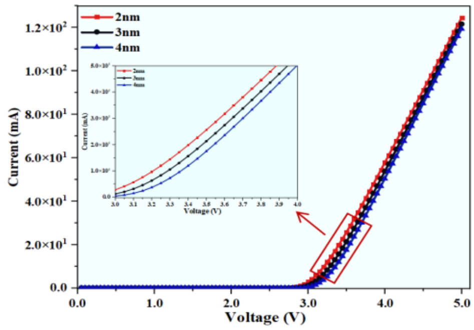


Figure 13 VI characteristic Curve of GaN-based LED with different Quantum well thickness

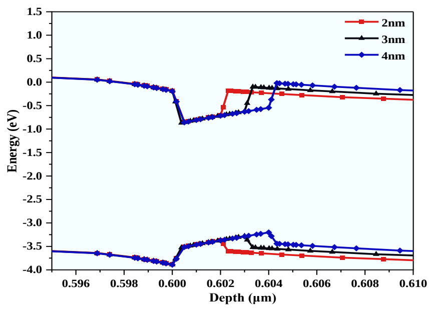


Figure 14 Energy band structure of quantum wells with different thickness

This is due to the increase in the thickness of the quantum well, the enhancement of the polarization effect, the intensification of the QCSE and the decrease of the quantum efficiency of LED [24-26]. Figure 15 shows the variation of LED luminous power with injection current at different quantum well thicknesses. When the quantum well thickness decreases, the LED's luminous power increases.

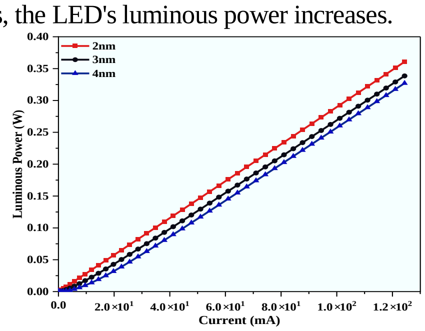


Figure 15 Variation of LED luminous Power with injection current under different Quantum well thickness

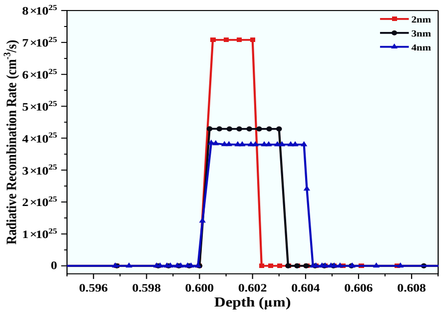


Figure 16 The curve of the radiative recombination efficiency of LED with the thickness of the quantum well

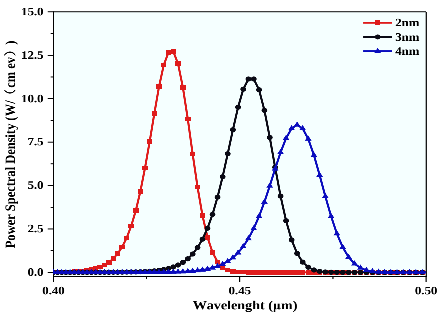


Figure 17 Optical Power density of GaN-based LED with different Quantum well thickness

Figure 16 shows the variation of the radiation recombination efficiency of LEDs with the thickness of the quantum well. The internal quantum efficiency declines as the quantum well's thickness increases due to a reduction in radiation recombination efficiency, this also validates the results in figure 17. The luminous power of LEDs decreases with the increase in thickness of the quantum well, and the peak wavelength shifts to the long wavelength, which leads to the red shift [27]. We think that properly reducing the thickness of the quantum well will increase the luminous intensity of the device, the electric field formed by the carriers in the quantum well weakens the polarization electric field, reduces the influence of the QCSE, and the effective band gap of the quantum well expands obviously, which leads to the blue shift of the LED spectrum [28-29].

3 Conclusion

This article primarily simulates the forward voltage, the content of In in the quantum well, and the quantum well thickness. The results show that with the increase in voltage, the influence of QCSE is weakened, the peak wavelength of LEDs decreases gradually, and the luminescence spectrum shows an obvious blue shift. The range of In components set is 18% to 30%. the increase of the content of in will aggravate the polarization effect, cause the energy band to tilt, the injection of carriers is blocked, and the radiation recombination efficiency is reduced, so the luminous efficiency decreases, and the peak value of the spectrum shifts to the long wave, that is, the redshift occurs. As the thickness of the quantum well is reduced, the electric field created by the carrier weakens the polarization electric field, increasing the effective band gap of the quantum well, shifting the peak wavelength to the short wave, and resulting in a blue shift in the luminescence spectrum.To sum up, for the GaN-based blue LED with this structure, when the well thickness of the InGaN/GaN quantum well is set to 3 nm and the In content is 20%, the device shows good optoelectronic properties. This paper has a basic reference value for the actual design, optimization and related research of GaN-based blue LEDs.

References

[1] Jinmin Li, Junxi Wang, Xiaoyan Yi, et al.. "IIINitrides Light Emitting Diodes: Technology and Applications"[D], Springer Science and Business Media LLC, 2020.

[2] Hai Wang, Le Wang, Jie Sun, et al.. Role of surface microstructure and shape on light extraction efficiency enhancement of GaN micro-LEDs: A numerical simulation study[J]. Displays, 2022(73): 102172.

[3] Yang himin, Li, Wangyang, Chen Xiaohang, et al.. Maximum efficiency and parametric optimum selection of a concentrated solar spectrum splitting photovoltaic cell-thermoelectric generator system[J]. ENERGY CONVERSION AND MANAGEMENT, 2018, 174(1): 65-71.

[4] Jia Chuanyu, He Chenguan, Wang Qi, et al.. Performance improvement of InGaN LEDs by using strain compensated last quantum barrier and electron blocking layer[J]. Optik, 2021(2):248-249.
[5] Haixia Wang,Honghao Liu,Huayu Zhang, et al.. A New Method for Bearing Steel Ball Surface Detection with Eddy Current Sensor[J]. Intelligent Robotics and Applications, 2021(13013): 263-274.
[6] Mannargudi, Anand, Vasileska, et al.. Quantum confinements in highly asymmetric sub-micrometer device structures[J]. Superlattices and microstructures, 2003(34): 347-354.
[7] Hui Wang, Xiaohong Yang, Rui Wang, et al.. Low Dark Current and High Gain-Bandwidth Product of Avalanche Photodiodes: Optimization and Realization [J]. Optics express, 2020, 28(11): 16211-16229.
[8] Rongier C, Gilblas R, Le Maoult Y, et al.. nfrared thermography applied to the validation of thermal simulation of high luminance LED used in automotive front lighting. [J]. Infrared Physics & Technology, 2022(4):120.
[9] Salhi A, Hadfield A, Muttlak S, et al.. Design and analysis of GaAs/AlAs asymmetric spacer layer tunnel diodes for high-frequency detection.[J]. Physica E, 2021(130): 114723.
[10] Rashid, Muhammad Haroon, Koel Ants, et al.. Cheema, Salman Arif.Modeling and simulations of 4H-SiC/6H-SiC/4H-SiC single quantum-well light emitting diode using diffusion bonding technique[J]. Micromachines, 2021,12(12): 1499.
[11] Bouchachia Badia;Hamdoune, Abdelkader. Improvement of InGaN/GaN Blue LED Performance by a BGaN Back-Barrier Layer: Simulation Study[J]. Journal of Nanoelectronics and Optoelectronics, 2019, 14(2): 147-153.
[12] Nour El I. Boukortt;Mabrouk Adouane;Rawan AlHammadi.High-performance ultrathin Cu(In,Ga)Se2 solar cell optimized by silvaco tools[J]. Solar Energy, 2021(228): 282-289.
[13] Liu Yibo, Wei Feng, Lu Tao, Liu, Zhaojun, et al.. GaN HEMT-LED Homogeneous Integration for Micro-LED Mass Transferring [J]. SID Symposium Digest of Technical Papers, 2018,(49): 660-664.
[14] Ahmad, Habib;Hussain, Shahzad;Shah, et al.. TCAD design of InGaN-based monolithic multi-wavelength LED with controlled Power spectral distributions[J]. Optik: Zeitschrift fur Licht- und Elektronenoptik: Journal for Light-and Electronoptic, 2015,126(21): 3140-3144.
[15] Luo Jiansheng, Chai Guangyue, Liu Wen, et al.. Optimization design of the chip structure for current distribution in the GaN-Based LED[J]. Semiconductor Technology, 2016,41(11): 822-826,880.
[16] N Boukortt, S Patane, B Hadri, Development of High-Efficiency PERC Solar Cells Using Atlas Silvaco(Article)[J]. Silicon, 2019,11(1): 145-152.
[17] Asma Belaid, Abdelkader Hamdoune.Numerical simulation of UV LEDs with GaN and BGaN single quantum well[J]. Journal of Semiconductors, 2019, 40(3): 47-52.
[18] Jia Chuanyu, He Chenguang, Liang Zhiwen, et al.. Improvement of Radiative Recombination Rate and Efficiency Droop of InGaN Light Emitting Diodes with In-Component-Graded InGaN Barrier.[J]. Physica Status Solidi. A: Applications & Materials Science, 2021,218(20): 1-8.
[19] Bouchachia, Badia;Hamdoune, Abdelkader. Improvement of InGaN/GaN Blue LED Performance by a BGaN Back-Barrier Layer: Simulation Study[J]. JOURNAL OF NANOELECTRONICS AND OPTOELECTRONICS, 2019,14(2): 147-153.
[20] Nour El I, Boukortt, Mabrouk Adouane, et al.. High-performance ultrathin Cu(In,Ga)Se2 solar cell optimized by silvaco tools[J]. Solar Energy, 2021(228): 282-289.
[21] Liu Yibo, Wei Feng, Lu Tao, et al.. GaN HEMT-LED Homogeneous Integration for Micro-LED Mass Transferring[J]. SID Symposium Digest of Technical Papers, 2018(49): 660-664.
[22] Ahmad, Habib;Hussain, Shahzad;Shah, Ijteba-ul-Hasnain.TCAD design of InGaN-based monolithic multi-wavelength LED with controlled Power spectral distributions[J]. Optik: Zeitschrift fur Licht- und Elektronenoptik: = Journal for Light-and Electronoptic, 2015, 126(21): 3140-3144.
[23] Simon Kaiser, Franz Symalla, Tobias Neumann, et al.. Automated Multiscale Design Flow for Organic LED Design[J]. ECS Meeting Abstracts, 2021(01): 1059.
[24] Luo Jiansheng, Chai Guangyue, Liu Wen, et al.. Optimization design of the chip structure for current distribution in the GaN-Based LED[J]. Semiconductor Technology, 2016, 41(11): 822-826.
[25] Benslim Amina, Meftah Amjad M, Labed Madani, et al.. Study and optimization of InGaN Schottky solar cell performance[J]. Optik, 2021, 247.
[26] Adaine A, Hamady S O S, Fressengeas N, et al.. Simulation study of a new InGaN p-layer free Schottky based solar cell[J]. Superlattices and Microstructures, 2016(96): 121-133.
[27] Abdoulwahab Adaine, Sidi Ould Saad Hamady. Nicolas Fressengeas.Simulation study of a new InGaN p-layer free Schottky based solar cell[J]. Scientific Research, 2016(11):38-41.
[28] Alam S, Sundaram S, Haas H, et al.. Investigation of p-contact performance for indium rich InGaN based light emitting diodes and solar cells[J]. PHYSICA STATUS SOLIDI A-APPLICATIONS AND MATERIALS SCIENCE, 2017,214(4): 1600496.
[29] Abdoulwahab Adaine, Sidi Ould Saad Hamady. Nicolas Fressengeas.InGaN Metal-IN Solar Cell: optimized efficiency and fabrication tolerance[J]. Scientific Research, 2017(11):65-70.

Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12719

Research Article Open Access

Application of SABO-VMD-KELM in Fault Diagnosis of Wind Turbines

Yuling HE[*], Hao CUI

Department of Mechanical Engineering, North China Electric Power University, Baoding, Hebei, 071003, China

[*]***Corresponding Author:*** *Yuling HE, E-mail: heyuling1@163.com*

Abstract

In order to improve the accuracy of wind turbine fault diagnosis, a wind turbine fault diagnosis method based on Subtraction-Average-Based Optimizer (SABO) optimized Variational Mode Decomposition (VMD) and Kernel Extreme Learning Machine (KELM) is proposed. Firstly, the SABO algorithm was used to optimize the VMD parameters and decompose the original signal to obtain the best modal components, and then the nine features were calculated to obtain the feature vectors. Secondly, the SABO algorithm was used to optimize the KELM parameters, and the training set and the test set were divided according to different proportions. The results were compared with the optimized model without SABO algorithm. The experimental results show that the fault diagnosis method of wind turbine based on SABO-VMD-KELM model can achieve fault diagnosis quickly and effectively, and has higher accuracy.

Keywords: *Wind turbine generator; Fault diagnosis; Subtraction-Average-Based Optimizer (SABO);Variational Mode Decomposition (VMD); Kernel Extreme Learning Machine (KELM)*

1 Introduction

Wind energy has the characteristics of clean, environmental protection, large reserves and wide distribution [1]. It has become one of the indispensable energy sources in today's world, and wind turbines are also all over the world. Due to its poor working conditions and complex load conditions, in order to ensure the normal operation of the wind turbine and reduce the cost of operation and maintenance, it is necessary to overhaul the wind turbine regularly. Wind turbine fault types can be divided into electrical faults and mechanical faults, and air gap eccentricity [2] and inter-turn short circuit faults [3] are the fault types with high frequency in mechanical and electrical faults, respectively. Wind turbine fault will have a significant impact on safe and efficient production, so the research on fault diagnosis method of wind turbine is of great significance to ensure the steady operation of wind turbine and equipment maintenance.

The generator fault signal is often accompanied by periodic impact, showing nonlinear and non-stationary characteristics. A large number of scholars have studied its fault diagnosis method [4-7]. In literature [8], the deep convolutional network is used to extract the vibration signal features of wind turbine, and then the fault classification is completed through the fully connected neural network. The results show that the fault diagnosis rate of this method is higher than that of other comparison methods. Jing Huang and Ruping Lin [9] et al. used whale optimization algorithm to optimize variational mode decomposition (VMD) with sample entropy as their fitness function, and then extracted the optimal intrinsic mode functions (IMFs) energy entropy to realize the generator inter-turn short circuit fault diagnosis, which improved the accuracy of fault diagnosis. Kernel Extreme Learning machine (KELM) is an extension of extreme learning machine, which can deal with nonlinear problems efficiently. Iiterature [10] use the particle swarm optimization kernel extreme learning machine to diagnose the fault of rotating machinery, and a modified hierarchical multi-scale dispersion entropy calculation method is proposed. The results show that this methods can well complete the fault diagnosis of rotating machinery. The above scholars have proved the effectiveness of VMD and KELM in generator signal processing, but the selection of VMD and KELM parameters has a great impact on their result [11]. Yong Lv [12] proposed a VMD optimization algorithm based on variable bandwidth control parameter strategy and center frequency adaptive convergence strategy. The actual case verifies that the optimized VMD can obtain more accurate and efficient results.

Received on October 3, 2023; Accepted on November 21, 2023

Pavel Trojovsky and Mohammad Dehghani proposed the Subtraction-Average-Based Optimizer (SABO) [13] algorithm which is a mathematical metaheuristic algorithm based on mathematical concepts, foundations and operations in 2023. This method has fast convergence speed and good optimization results. This can prevent the dependence of the algorithm on specific middle group members, and through the optimization of this algorithm, it can avoid falling into local optimal solutions.

This paper proposes a wind turbine fault diagnosis method based on SABO-VMD-KELM model.Firstly, the SABO algorithm was used to optimize the VMD parameters to obtain the best modal components, and the features of the modal components were calculated to form a feature vector, and the training set and the test set were divided according to different proportions. Secondly, the SABO algorithm was used to optimize the KELM parameters by using the error rate of the training set and the test set as the fitness function, and the SABO-KELM model was established. Finally, the test set data is input for fault diagnosis and classification.

2 Fault Diagnosis Methods

2.1 Basic Principle of SABO

Unlike group-based metaheuristic algorithms inspired by various natural collective phenomena, the SABO algorithm is a mathematical metaheuristic algorithm based on mathematical concepts, foundations, and operations. This method features fast convergence and effective optimization results. The specific principles of the SABO algorithm are as follows:

(1) The search agent positions are randomly initialized in the search space.

$$x_{i,d} = lb_d + r_{i,d} \cdot (ub_d - lb_d) \quad i = 1,\cdots,M \quad d = 1,\cdots,n \tag{1}$$

$x_{i,d}$ is its dth dimension in the search space (decision variable), M is the number of search agents, n is the number of decision variables, $r_{i,d}$ is a random number between 0 and 1, and lb_d and ub_d are the lower and upper bounds of the dth decision variable.

(2) The iterative process of SABO.

The SABO algorithm is based on a newly proposed operation " $\bar{}v$",which is defined by the −subtraction of the search agents Q from the search agent P,which is shown as follows:

$$P \, -_{\upsilon} \, Q = \text{sign}\big(F(P) - F(Q)\big)\big(P - \vec{\upsilon} * Q\big) \tag{2}$$

where $\bar{}\,v$ is a vector of the dimension n, in which components are random numbers that are generated between 0 and 1, the operation "∗" represents the Hadamard product of the two vectors, F(P) and F(Q) are the values of the objective function of the search agents P and Q.

In the SABO method we've proposed, the movement of a given search agent X_i within the exploration space is determined by averaging the differences between every other search agent X_j (where j equals 1, 2, ..., M) and the search agent X_i. Therefore, each search agent's updated location is computed using equation (3).

$$X_i^{new} = X_i + \vec{r}_i * \frac{1}{N}\sum_{j=1}^{M}\left(X_i -_{\upsilon} X_j\right) \quad i = 1, 2\cdots, M \tag{3}$$

where X_i^{new} is the new proposed position for the ith search agent X_i , M is the total number of the search agents, and $\vec{r}_i$ is a vector of the dimension n, in which components have a normal distribution with the values from the interval [0, 1].

Then, if this proposed new position leads to an improvement in the value of the objective function, it is acceptable as the new position of the corresponding agent, according to (4).

$$X_i = \begin{cases} X_i^{new} & F_i^{new} < F_i \\ X_i & else \end{cases} \tag{4}$$

where F_i and F_i^{new} are the objective function values of the search agents X_i and X_i^{new}, respectively.

2.2 VMD basic principles and optimization methods

Variational Mode Decomposition is an adaptive signal processing algorithm for handling non-stationary and nonlinear signals. It can determine the central frequency and bandwidth of each intrinsic mode function component (IMFs) of the signal under a variational constraint framework, with the total bandwidth being minimal. The number of components is determined by the predefined number of decomposition levels K, then transforming the decomposition process of the original signal into the solving process of a variational problem. The specific steps are as follows:

(1) obtain the unilateral spectrum of IMFs.

$$A = \left[\delta(t) + \mathrm{j}/\pi t\right] * u_k(t) \tag{5}$$

Where $k = 1, 2, \cdots, K$, $\delta(t)$ is the impulse function, and $u_k(t)$ is the k-th IMF obtained from decomposition.

(2) Adjust the IMF spectrum to its fundamental band.

Introducing the exponential operator $\mathrm{e}^{-\mathrm{j}\omega_k t}$ to adjust the spectra of each modal component to their fundamental frequency band B.

$$B = \left[\left(\delta(t) + \mathrm{j}/\pi t\right) * u_k(t)\right]\mathrm{e}^{-\mathrm{j}\omega_k t} \tag{6}$$

Where ω_k is the center frequency of the kth IMF.

(3) Demodulate each IMF, estimate the bandwidth of each modal component, and construct a constrained variational model.

$$\begin{cases} \min\limits_{(u_k,\,\omega_k)} \left\{ \sum\limits_{k=1}^{K} \left\| \partial t\left[\left(\delta(t) + \mathrm{j}/\pi t\right) * u_k(t)\right]\mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \; P(t) = \sum\limits_{k=1}^{K} u_k \end{cases} \tag{7}$$

Where $\delta(t)$ is the Dirac distribution function and $\| \quad \|_2$ is the 2-norm.

To solve the above constraint model, the quadratic penalty factor and Lagrange multiplier are introduced, so that the constraint problem can be transformed into an unconstrained problem.

$$L(u_k,\omega_k,\lambda)= \alpha\sum_{k=1}^{K}\left(\left\| \partial t\left[\left(\delta(t)+\mathrm{j}/\pi t\right)*u_k(t)\right]\mathrm{e}^{-\mathrm{j}\omega_k t}\right\|_2^2\right)+ \left\|P(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2+\left\langle\lambda(t),P(t)-\sum_{k=1}^{K}u_k(t)\right\rangle \tag{8}$$

Where α is the quadratic penalty factor, and 〈 〉 represents the inner product between vectors.

The central frequency method [14] is usually used to determine the number of modal components in VMD. In this paper, the SABO algorithm is used to optimize the VMD's secondary penalty factor and the number of modal components K. The paper selects the minimum envelope entropy as the fitness function for VMD. Envelope entropy represents the sparsity characteristic of the original signal; when an IMF contains more noise and less characteristic information, the envelope entropy is higher, and vice versa. The fitness function serves as a criterion in the SABO-optimized VMD algorithm for evaluating the quality of search agent positions and is also an important factor influencing the quality of VMD results. The other parts of this paper are arranged as follows. Section II introduces the principle of fault diagnosis method, Section III summarizes the method process, and finally, the conclusion of this paper is obtained in the Section IV.

2.3 SABO-KELM

Kernel Extreme Learning Machine (KELM) is an improved algorithm based on Extreme Learning Machine (ELM) and optimized through a kernel function. KELM is capable of enhancing the predictive performance of the model while retaining the advantages of ELM. ELM is a single hidden layer feedforward neural network, whose objective function represented in matrix form is as follows.

$$F(x)=h(x)\beta=L \tag{9}$$

Where x is the input vector,h(x) and H are the hidden layer node outputs, β is the output weight; L is the desired output.

Transform the network training into a problem of solving a linear system. It is determined according to the formula, where is the generalized inverse matrix of H. Introduce the regularization coefficient C and the unit matrix I to enhance the stability of the neural network. The least squares solution of β is shown in the formula(10-12).

$$\beta=H^*L \tag{10}$$

$$H^*=H^T\left(HH^T\right)^{-1} \tag{11}$$

$$\beta=H^T\left(\frac{I}{C}+HH^T\right)^{-1}L \tag{12}$$

In this paper, Radial-Basis-Kernel function is introduced into ELM, and the kernel matrix is shown in formula(13-15), where is the formula of Radial-Basis-Kernel function.

$$K(x_i,x_j)=\exp\left(-\frac{\left\|x_i-x_j\right\|^2}{\delta^2}\right) \tag{13}$$

$$\Omega_{KELM}=HH^T=h(x_i)h(x_j)=K(x_i,x_j) \tag{14}$$

The objective function can be expressed as:

$$F(x)=\begin{bmatrix}K(x,x_1)\\ \vdots \\ K(x,x_N)\end{bmatrix}\left(\frac{I}{C}+\Omega_{KELM}\right)^{-1}L \tag{15}$$

Where $x_1,\cdots,x_N$ is the given training sample, N is the number of samples.

This paper optimizes the regularization coefficient C and kernel coefficient of KELM using the SABO algorithm. The paper selects the error rate of the training and test sets as the fitness function for SABO-KELM.

3 Fault Diagnosis Process

Since the selection of parameters for VMD and KELM can impact their results, this paper uses the SABO algorithm to optimize the VMD secondary penalty factor、the number of modal components and the KELM kernel coefficient、 regularization coefficient.

The main method process of this paper is shown in the figure, and the specific steps are as follows:

(1) Decompose the original signal using VMD;

(2) Set the number of iterations, optimization dimensions, and optimization bounds for the SABO algorithm, using the minimum envelope entropy as the fitness function, and apply the SABO algorithm to optimize the number of VMD modal components and the secondary penalty factor;

(3) Obtain the optimal IMF and calculate its mean, variance, kurtosis, effective value, peak factor, impulse factor, dispersion entropy, permutation entropy, and sample entropy to construct a feature vector;

(4) Set parameters for the SABO algorithm, and use the error rate of the training and test sets to optimize the regularization coefficient C and kernel coefficient of the KELM model;

(5) Select the Radial-Basis-Kernel function as the kernel function for KELM, input the training set feature matrix to train the KELM model;

(6) Input the test set data into the SABO-KELM model to obtain prediction results.

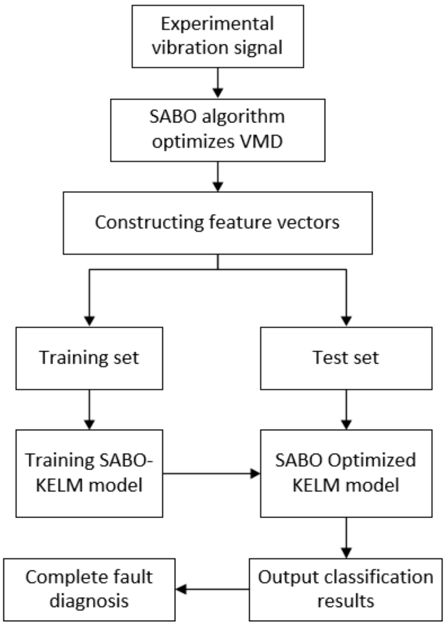


Figure 1 Flowchart of fault diagnosis

4 Experimental Analysis

4.1 Experimental signal acquisition

To verify the effectiveness of the proposed SABO-VMD-KELM model in this paper, experimental data from a LR-5 permanent magnet generator at North China Electric Power University is used. The test unit is an inner rotor permanent magnet wind power generator with a rated power of 5KW and a sampling frequency of 5000Hz. Simulations were carried out for four different states of the generator under full load operation: normal operation, 10% radial static eccentricity fault of the air gap (RSAGE), 10% short-circuit fault in stator windings between phases A1-A4 (SISC), and two combined faults (RSAGE & RISC). Vibration signals from the stator were collected using a vibration acceleration sensor. The table below shows some of the collected data.

Figure 2 Wind turbine experimental setup

4.2 Analysis of VMD results

Using SABO to optimize the VMD decomposition of stator vibration signals, obtaining the optimal IMF based on the minimum envelope entropy as the fitness function. Calculate its mean, variance, kurtosis, effective value, peak factor, impulse factor, dispersion entropy, permutation entropy, and sample entropy to construct a feature vector. Choose 100 signal groups under each of the four conditions – normal, RSAGE, SISC, RSAGE&RISC (labeled as 1, 2, 3, 4, respectively) – totaling 400 signal groups as input and label them as shown in the table below. Due to space limitations, the table only lists the effective value from the feature vector. Figure 3 shows the distribution of the effective values of the best IMF components in each group, which can only roughly distinguish the four operating states of the generator, but cannot achieve accurate diagnosis, so it needs to be further imported into KELM for analysis.

Table 1 Grouping eature vectors

Serial Number	Valid value	Lable
1	2.5439	1
…	…	…
100	3.5242	1
101	4.5712	2
…	…	…
200	4.5803	2
201	3.2327	3
…	…	…
300	2.7084	3
301	2.6484	4
…	…	…
400	2.0802	4

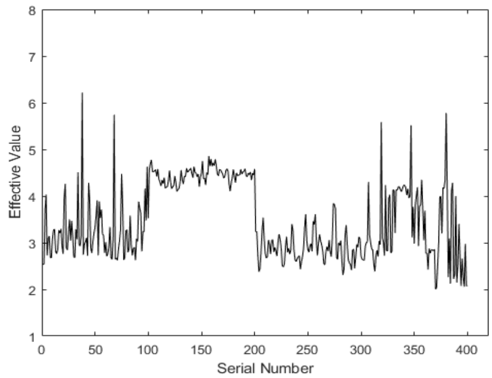


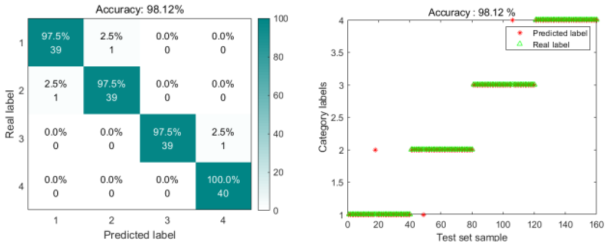
Figure 3 Effective value distribution of the best IMF components for each group

4.3 Analysis of KELM results

To validate the effectiveness of the SABO-VMD-KELM model in fault identification and the impact of the training set proportion on the results, training and test sets with different data partition ratios were randomly extracted. The test set was then input into the model for classification, and the results are presented in the table. The results demonstrate that as the proportion of the training set gradually increases, the accuracy of the model's recognition also improves. This proves the good stability of the SABO-VMD-KELM model, which does not suffer from overfitting or underfitting, though the model training time also increases. The figure below shows the confusion matrix when the training set accounts for 60%.

Table 2 Results for different test set proportions

Percentage of training set	30%	40%	50%	60%
Accuracy/%	93.2	95.6	96.8	98.1
Testing time/s	3.2	3.8	4.5	5.1



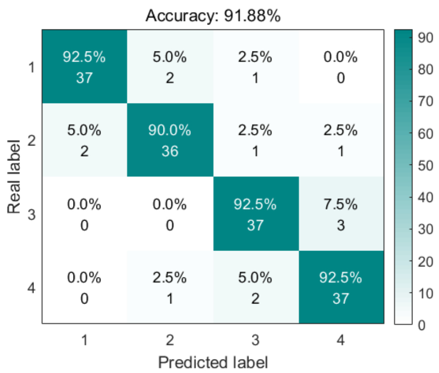
(a) (b)

Figure 4 Fault diagnosis results

4.4 Comparison of methods

To further illustrate the effectiveness of the method in this paper, The SABO-KELM model is compared with the particle swarm optimization (PSO) KELM model, SABO-optimized VMD feature vectors were imported into SABO-KELM and PSO-KELM models, and the fault accuracy rates were calculated as shown in the figure. The table shows that SABO-VMD-KELM outperforms PSO-KELM in terms of accuracy for normal, RSAGE, SISC, and RSAGE&RISC conditions, and also has a shorter training time, demonstrating the effectiveness and speed of the method presented in this paper.

(a)SABO-KELM



(b)PSO-KELM

Figure 5 Comparison of method results

Both models were subjected to ten repeated experiments, and the results showed that the correct diagnosis rate of SABO-VMD-KELM was consistently higher than that of PSO-KELM in all ten experiments.

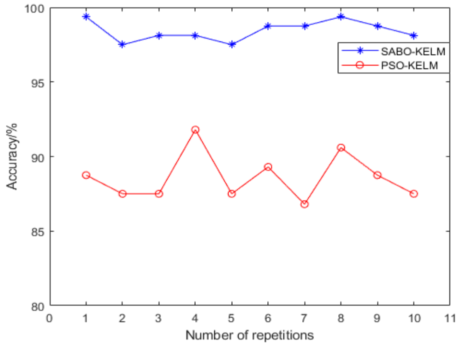


Figure 6 The results of ten times

Table 3 Accuracy comparison

Method	Average accuracy/%
SABO -KELM	98.75
PSO-KELM	91.88

5 Case Study of External Rotor Permanent Magnet Generator

5.1 Experimental signal acquisition

In order to further verify the effectiveness of the proposed method, the author uses the data of the outer rotor permanent magnet power generation motor model test unit for analysis. The following figure shows the experimental unit diagram of the outer rotor permanent magnet power motor model.

Figure 6 Outer rotor permanent magnet generator set

The moving-mode unit can simulate the external rotor permanent magnet generator RAGE, SISC, RSAGE&SISC, simulate the operation of external rotor permanent magnet generator by driving motor, and collect the vibration data of generator stator with sampling frequency of 5000Hz. The author recorded four kinds of generator fault data: normal, RSAGE, RISC, and RSAGE&SISC. The vibration data of each state was divided into 100 groups of signal sequences with length of 2048, a total of 400 groups of samples, and the proportion of training set and test set was 60%. The specific data information description is shown in the table:

Table 4 Data information description

Type of fault	Number of training set groups	Lable
NORMAL	40	1
RSAGE	40	2
SISC	40	3
RSAGE&SISC	40	4

5.2 Analysis of results

The author used SABO algorithm to optimize VMD parameters to extract the best IMF effective values for each group, and the results are shown in Figure 8.

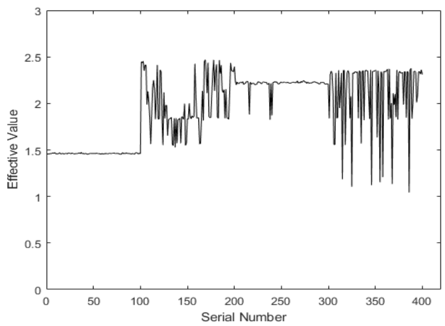
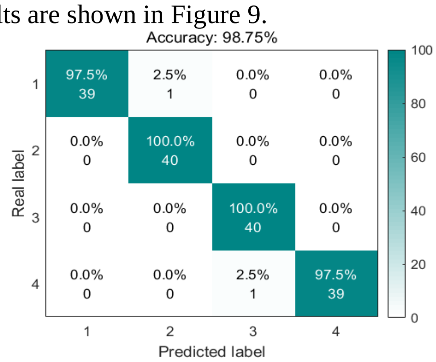
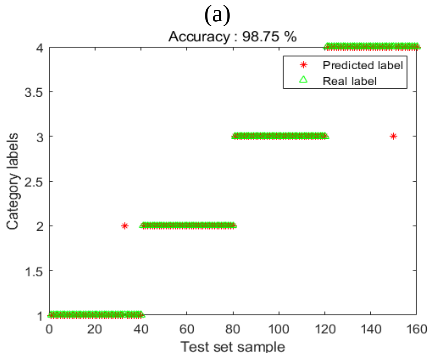


Figure 8 Effective value distribution of the best IMF components for each group

The characteristics of different faults in Figure 3 are slightly different and have a certain degree of identification, but precise fault diagnosis and classification cannot be achieved. Therefore, the extracted feature vectors are input into the SABO-KELM model for training and testing to obtain the fault diagnosis rate, and the results are shown in Figure 9.



(a)



(b)

Figure 9 Fault diagnosis results

Only 2 of 160 groups of data were diagnosed incorrectly, and the accuracy of fault diagnosis reached 98.75%, which proved the effectiveness of the proposed method again.

6 Conclusion

In order to improve the fault diagnosis accuracy of wind turbine, this paper proposes a wind turbine fault diagnosis method based on SABO-VMD-KELM model. The conclusions are as follows:

(1) The SABO algorithm was applied to optimize the parameters of VMD and KELM, and the resulting SABO-VMD-KELM model achieved an average fault diagnosis accuracy rate of 98% for wind turbines. This demonstrates the effectiveness of the model in diagnosing faults in wind turbines and provides a reference method for research in fault diagnosis of wind turbines.

(2) Using the same SABO-optimized VMD to decompose the original signal data, the accuracy rate of SABO-KELM is higher than that of PSO-KELM, demonstrating that SABO-KELM has a clear advantage in terms of generalization performance and accuracy.

References

[1] H. Polinder, J. A. Ferreira, B. B. Jensen, et al.. Trends in Wind Turbine Generator Systems [J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2013,1(3):174-185.

[2] Yu-Ling H. Impact of Static Air-Gap Eccentricity on Thermal Responses of Stator Winding Insulation in Synchronous Generators[J]. IEEE Transactions on Industrial Electronics, 2022,69(12):13544-13554.

[3] Yu-Ling H. Impact of Stator Interturn Short Circuit Position on End Winding Vibration in Synchronous Generators [J]. IEEE Transactions on Energy Conversion, 2022,36(2):13544-13554.

[4] M. Khov, J. Regnier and J. Faucher. Detection of turn short-circuit faults in stator of PMSM by on-line parameter estimation [C]. 2008 International Symposium on Power Electronics, Electrical Drives, Automation and Motion, 2008.

[5] Z. Gao, S. X. Ding and C. Cecati. Real-time fault diagnosis and fault-tolerant control [J]. IEEE Transactions on Industrial Electronics, 2015,62(6):3752-3756.

[6] T. Goktas, M. Zafarani and B. Akin. Discernment of Broken Magnet and Static Eccentricity Faults in Permanent Magnet Synchronous Motors [J]. IEEE Transactions on Energy Conversion, 2016,31(2):578-587.

[7] J. -K. Park and J. Hur. Detection of Inter-Turn and Dynamic Eccentricity Faults Using Stator Current Frequency Pattern in IPM-Type BLDC Motors [J]. IEEE Transactions on Industrial Electronics, 2016,63(3):1771-1780.

[8] Y. Shen, B. Chen, F. Guo, et al.. A Modified Deep Convolutional Subdomain Adaptive Network Method for Fault Diagnosis of Wind Turbine Systems [J]. IEEE Transactions on Instrumentation and Measurement, 2022,71:1-10.

[9] J. Huang, R. Lin, Z. He, et al.. Application of WOA-VMD-SVM in Fault Diagnosis of Generator Inter-turn Short Circuit [C]. 2022 China Automation Congress (CAC),2022.

[10] F. Zhou, J. Shen, X. Yang, et al.. Modified Hierarchical Multiscale Dispersion Entropy and its Application to Fault Identification of Rotating Machinery [J]. IEEE Access, 2020,8: 61361-161376.

[11] X. Wang, G. Sui, J. Xiang, et al.. Multi-Domain Extreme Learning Machine for Bearing Failure Detection Based on Variational Modal Decomposition and Approximate Cyclic Correntropy [J]. IEEE Access, 2020,8: 97711-197729.

[12] Y. Lv, Z. Li, R. Yuan, et al.. Variable-Bandwidth Self-Convergent Variational Mode Decomposition and its Application to Fault Diagnosis of Rolling Bearing [J]. IEEE Transactions on Instrumentation and Measurement, 2023,73:1-15.

[13] Trojovský, P., Dehghani. Subtraction-Average-Based Optimizer: A New Swarm-Inspired Metaheuristic Algorithm for Solving Optimization Problems[J]. Biomimetics, 2023,8(1):149-152.

[14] C. Wang, H. Li, G. Huang, et al.. Early Fault Diagnosis for Planetary Gearbox Based on Adaptive Parameter Optimized VMD and Singular Kurtosis Difference Spectrum [J]. IEEE Access, 2019,7:31501-31516.

Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12720

Research Article Open Access

A System of Image Recognition-Based Railway Foreign Object Intrusion Monitoring Design

Beiyuan WANG, Lingqi WANG*, Chuanya GU

Shanghai Bearing Technology Research Institute Co., Ltd., Shanghai, 201801, China

***Corresponding Author:** Lingqi WANG, E-mail: 327611126@qq.com*

Abstract

The monitoring system designed in this paper is on account of YOLOv5 (You Only Look Once) to monitor foreign objects on railway tracks and can broadcast the monitoring information to the locomotive in real time. First, the general structure of the system is determined through demand analysis and feasibility analysis, the foreign object intrusion recognition algorithm is designed, and the data set required for foreign object intrusion recognition is made. Secondly, according to the functional demands, the system selects a suitable neural web, and the programming is reasonable. At last, the system is simulated to validate its functionality (identification and classification of track intrusion and determination of a safe operating zone).

Keywords: *Railway; Deep learning; YOLOv5; Image intelligent recognition; Obstacle detection*

1 Introduction

Object detection in YOLO is seen as a regression concern, which determines the position of a bounding box and works out the probability of it attributing to a particular category. Use a single neural network to predict bounding box and class probabilities directly from the complete image in one evaluation [1].

In the official code of YOLOv5 used in this project, the target detection networks given are respectively YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x models. YOLOv5 code network file is yaml format, network structure is very concise, using Leaky ReLU and Sigmoid as YOLOv5 activation function [2].

YOLOv5 passes each batch of training data through the data loader and simultaneously enhances the training data. The data loader performs three kinds of data enhancement: scaling, color space adjustment, and Mosaic enhancement (Mosaic data enhancement does effectively solve the most troublesome "small object problem" in model training, that is, small objects are not as accurately detected as large objects).

The authors of YOLOv5 provide us with two optimization functions, Adam, and SGD, and both have preset training hyperparameters to match (Default: SGD). The authors of YOLOv5 recommend that Adam is a more appropriate choice if you need to train smaller custom datasets, although Adam generally has a lower learning rate than SGD. But if you train large data sets, SGD works better for YOLOv5 than Adam. In fact, there has been no unified conclusion on which SGD or Adam is better in the academic community, depending on the actual project situation. If the optimization function is trained on a small custom data set, ADAM is a more suitable choice, but if the training of a large data set, the SGD optimization function is more frequently used.

2 Feasibility Analysis

2.1 Technical feasibility

The YOLOv5 network model is the underlying structure of the back end of the system. This system wants to achieve efficient and high-speed recognition, in that the YOLOv5m network model [3] is adopted for data training. The data set was processed and annotated using the LabelImg tool to obtain the normalized data needed for YOLOv5 training. To sum up, method used in this paper is simple and clear, and can be completed in a short time. Besides, the front- and back-end interactions are hierarchical, and technically feasible.

2.2 Operational feasibility

The system uses a simple framework, reasonable training mode and design language to ensure the order and efficiency in the development process. Users can quickly complete their desired results through a clear interface UI. Therefore, the system is feasible in operation.

Received on October 19, 2023; Accepted on November 27, 2023

2.3 Economic feasibility

The development cost is not high, since YOLOv5 is an open-source project on the Internet, it can be downloaded for free use. The back end uses the CPU for training. Users only need to access the system and upload pictures while let the system do the rest. In general, the development cost of this project is small, so it's economically feasible.

3 Design of the System

3.1 Foreign object recognition algorithm design

Based on YOLOv5 algorithm model, this paper realizes the monitoring function of railway foreign objects intrusion, therefore, by obtaining the best training model (. best-py) after training the data set through the YOLOv5 algorithm model is the key to this function, and then proceed to the next step of foreign objects category judgment. The program background can call the YOLOv5 algorithm model, and the PyQt5 module in Python is used to realize railway foreign object monitoring interface [4].

After the training of YOLOv5 algorithm model is completed, the training file will be output to the user. User can set the format of general input, usually used .jpg format. The system will monitor the input information by invoking the training file. After monitoring, the interface will display the category size and influence range of foreign objects invaded by railway.

3.2 General structure of system

With the continuous progress of artificial intelligent machine learning, the Chinese rail transit industry can cut the cost of monitoring the railway. Due to the gradual emergence of image recognition assisted artificial monitoring, the vital research is to solve the security and dependability of the train undertaking.

The biggest threat the train faces in the process is the impact of foreign objects on the railway, and the impact is mainly caused by human, animals, or stones etc. on the railway, hindering the normal train speed conditions. To avoid similar accidents, the railway safety department will set up warning signs or protective measures such as guardrail accordingly, but it's impossible to strictly inspect or clean every area all the time where the train travels, and because the standard rate of the train is generally above 150km/h, the operation of deceleration or emergency braking cannot be carried out in time when the railway foreign objects in the line of sight is encountered. Therefore, it's crucial to design a railway foreign object intrusion monitoring system reminding train drivers to implement different train speeds when encountering railway foreign objects at different distances.

The monitoring process's interface of the entire railway foreign object intrusion system is carried out according to the overall system flowchart as indicated in Figure 1.

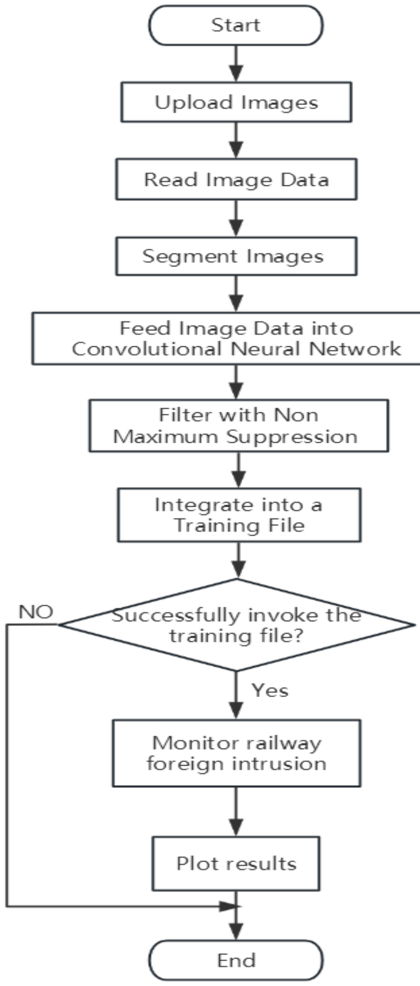


Figure 1 Flowchart of the system's overall structure

The specific process of this system is to upload the pictures that need to be monitored first and wait for the system to automatically read the images. The next step is to determine whether the system can successfully invoke the training file, which is trained previously. Therefore, a good recognition effect relates to the quality of the previously trained files. If the training file can be successfully called, the system will carry out the railway foreign object monitoring module, besides, it will return the foreign object position and affected region to the system after the checking is fulfilled. The system will classify different affected areas of the foreign object to different traveling speeds of the train according to the existing safe area range and transmit the data to the cab to give the driver the driving speed indication. Finish a series of commands to end the program of the system, exit the operation; If you cannot successfully call the training file, then directly end the program of the system, and exit the operation.

Based on the in-depth study of domestic and foreign railway foreign object intrusion monitoring technology, this subject intends to adopt YOLOv5 algorithm to conduct real-time monitoring of railway foreign objects intrusion penetration. Due to the constraints of the test environment, the experiment cannot be carried out while the train is running, so this subject will only analyze and discuss the theoretical test in the following chapters.

The system preparation work designed in this project: download the YOLOv5 project to the local, configure Python3.9, and the environment operating system is Windows 10; IDE: PyCharm; Python version: anaconda Python3.9. This system uses the CPU of this device for training. If large-scale training, image and video recognition and other operations are carried out, the computer should be equipped with GPU [5].

3.3 Creation of data set

Since, there are no training samples for railway foreign object intrusion monitoring in the existing known data sets, this topic intends to adopt the method of manual labeling to build a group of suitable data sets to ensure that the algorithm model can obtain good training results. With Baidu and other search engines as the main data set source, the web crawler developed by Python was used to collect and download related pictures. According to the actual railway operation, five types of foreign bodies are mainly monitored, including rock, person, animal, car, and rabbit. Integrate all image data and unify into.jpg format. A total of 324 data sets were used to train the algorithm model. After consulting the literature on image recognition, we can get a conclusion that the optimal data set required for computer image training is at least 300 pictures. Therefore, during the test, we plan to train more than 300 pictures first, and then continue to consider increasing the number of sample data sets if the effect is not as expected.

After sorting, the existing data sets need to be labeled to improve the performance of the algorithm model. LabelImg was used to manually mark 324 pictures in the foreign objects' image data set on the railway and label them with categories. The connection of LabelImg tag is indicated in Figure 2.

After manual annotation is completed, there are mainly two folders for storing related images and annotation information. One folder is called 'images and the other folder is called 'label'. The label folder stores the label information (in .txt format), that is, the label file for target detection. The label information corresponds to the name of the image data set file one by one.

Figure 2 The interface of LabelImg

Open the designated mark file, as shown in Figure 3 normalized data of railway foreign object categories and prior box coordinates. Each line in the .txt file represents a target, the x and y coordinates of the center point post-normalization, in addition to the breadth w and tallness h of the objective box. Besides, under the label folder, there is also a .txt file called 'classes', which contains information about the class names of all the label boxes. After the annotation of 324 images in the training data set, the pre-processing of the data set is completed.

0 0.382422 0.787500 0.757031 0.338889
1 0.417969 0.793750 0.031250 0.068056

Figure 3 Normalized data of railway foreign object categories and prior frame coordinates

3.4 Regional division function

Currently, two main methods of lane line detection exist domestically and internationally: one is a model-based approach, the other is a feature-based technique. The model-based detection method is to assign the lane to a sophisticated arithmetical model and fit the road line based on the model. The principle is to match the appropriate curve model for the lane lines according to the geometric features of the lane lines on the structured roadway, and then use the least square manner, Hough Transform to be the right size for lane principles [6]. The widespread mathematical representations are linear, allegorical and spline models. This method has strong anti-interference ability to noise, but it also has drawbacks, that is, one lane model cannot adapt to multiple road scenes at the same time.

Aspect-based detection method is to obtain the aspect information of lane lines from the pavement area through edge detection or threshold segmentation according to the feature information of lane lines themselves. This method has high requirements on the edge features of the lane lines and can obtain better results when the edge features are obvious, but it's very thoughtful to sound and has impoverished robustness.

This project adopts a Transmission Hove based road segment recognition algorithm, that is, a model-based recognition method to complete the division of this area. Figures 4 and 5 are the original diagram and the effect diagram after adding the algorithm respectively.

Figure 4 Original diagram

Figure 5 Effect diagram after adding the algorithm

First, the picture is changed into HLS hue space. And then using the edge detection and threshold methods to detect the lane lines [7], the edge detection results are combined with the color detection results, and the L-channel is used to suppress the non-white information. According to the lane line information detected above, histogram method and sliding window method [8] are used to determine the exact position of the track line. The safe region algorithm can realize 4 vertices in the return region, that is, the two regions in Figure 3.5 are separated into the crimson region and the yellow region. The algorithm can return a total of 8 vertices, and then judge which area in the crimson and yellow area the foreign object belongs to according to the center coordinates of the object position recognized by YOLO, to realize the train deceleration or braking operation.

3.5 Foreign object recognition function

In the railway foreign object identification module, the user uploads the images that need to be detected, and the system stipulates that the images can only be uploaded in .jpg format. If the uploaded format doesn't belong to the specified format, it won't be uploaded. If the user isn't satisfied with the uploaded picture, or if there is an error in the uploaded picture, the picture can be cleared and then uploaded again. If the user confirms that the uploaded picture is correct, the system can detect the current picture, and the successful detection result is returned to the system. The widespread procedure is indicated in Figure 6 Flowchart of railway foreign object recognition module.

The specific process of this system for foreign object recognition function is that the background monitoring personnel first upload the image that needs to be monitored, and if the image is successfully uploaded, wait for the system to automatically read the image; If the upload fails, upload the image again. The next step is to determine the user's next response, which can be divided into two steps, either to start recognizing the image uploaded by the user, or to clearly exit the image. If the user performs the identification undertaking, the system affects regardless of the successful identification. When the identification is successful, the monitoring result will be returned to the display without ending the program. While the discovery is not triumphant, the program of the system is ended directly, and the operation is withdrawn.

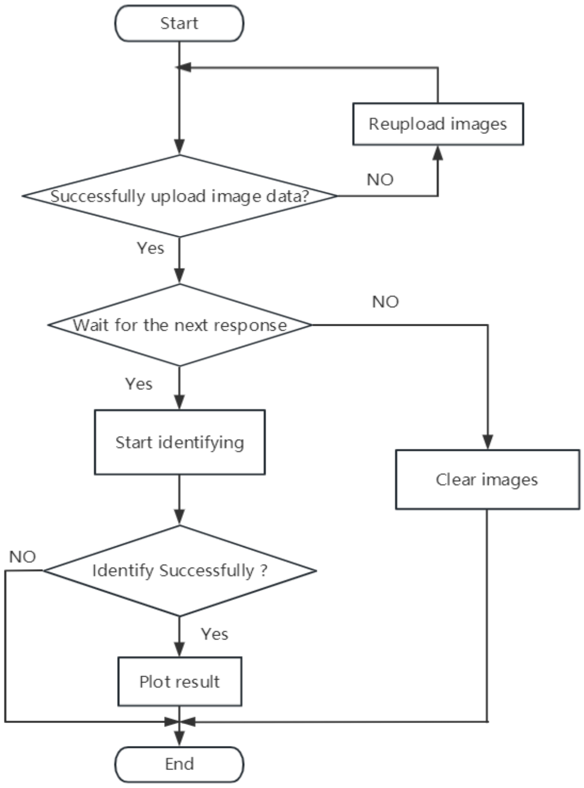


Figure 6 Railway foreign object identification module flowchart

Figure 7 shows the test recognition types and accurate results. It can be seen from the figure that the system has a high accuracy rate for the recognition effect of rail and a relatively low accuracy rate for the recognition effect of person, but the category judgment of both is correct. In view of the accuracy problem, in terms of subsequent training and data set collection, emphasis should be placed on the training and testing of 5 types of foreign bodies required for this subject (human, animal, stone, rabbit, car).

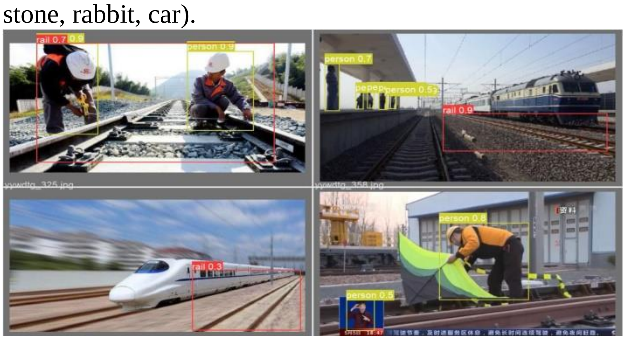


Figure 7 Recognition type and accuracy (partial)

4 Test and Analysis

4.1 Model training

YOLO algorithm is mainly composed of two parts, one is classification, the other is detection. The first method uses the image network for pre-training, while the second one uses precise settings on the detection data. YOLOv5 pre-training is based on COCO data set, its official pre-training weight is based on FP16, and the training weight is based on FP32, which can double the storage space. In common object detection algorithms, the image is uniformly reduced to a standard size (such as 416×416,608×608, commonly used by the YOLO algorithm) and transmitted to the detection network. However, YOLOv5 has been improved: In the actual use of the project, because various images, the size of each side is different, so after magnification, the size of the black edge on both sides is also discrepant, but more information will be added, thus affecting the speed of reasoning. In this way, on the original drawing, the adaptive addition of the smallest black edge can increase the reasoning speed by 37%.

When YOLOv5 code is trained, first modify the training configuration. To create a .yaml file, it's necessary to create a mask_data.yaml file in the data folder primarily, then to modify the address and category parameters of this file. The category name is changed to the data set label name used in this project (rail, animal, person, rock, car, rabbit), that is, the arrangement folder of the data set is modified, as indicated in Figure 8, and training can be carried out after modification.

```
# train and val data as 1) directory: path/images/, 2) file: path/images.txt, or 3)
train: D:/360Downloads/lunwen/ImgSpider/YOLO_Rail/score/images/train
val: D:/360Downloads/lunwen/ImgSpider/YOLO_Rail/score/images/val

# number of classes
nc: 6

# class names
names: ['rail','animal','person','rock','car','rabbit']
```

Figure 8 Modify the configuration file of the data set

4.2 Test result

Before model training, the labeling of sample data sets was divided into two days, and continuous labeling was not done, which affected the recognition of some classes in some labeling information, resulting in a certain deviation in the final picture identification of alien thing types. The confusion matrix of the training model is indicated in Figure 9.

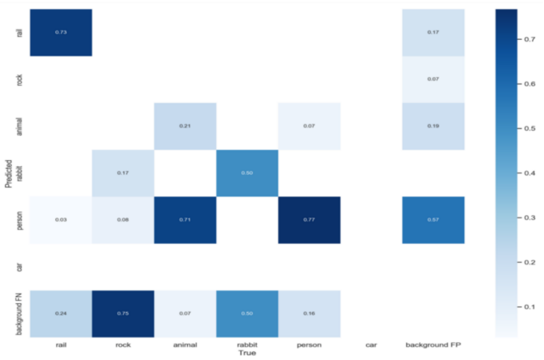


Figure 9 The confusion matrix of the training model

The explanation is as follows: the x-coordinate is the correct classification, and the y-coordinate is the classification predicted by the model.

The target detection model usually outputs many detection boxes, and we measure the detection effect of the model by counting and calculating whether each detection box can detect various proportions of the target. Therefore, we will divide the detection box into the following 4 cases, and the basis for judgment is mainly through calculating the intersection ratio IoU. The IoU intersection ratio function can be used to evaluate object detection algorithms [9]. By calculating the ratio of intersection and union of two bounding boxes, it can be used to determine whether the object detection algorithm is running well. The two compelled container are compelled box and ground reality reference standard respectively. Therefore, the IoU is designed to evaluate whether the object positioning algorithm is precise, that is, the IoU measures the relative size of the overlap between two compelled boxes. A way to measure positioning accuracy: As a general convention for computer detection, if the IoU is above 0.5, the prediction is acceptable; If the IoU equals to 1, the predicted and actual boxes overlap perfectly. The manual threshold for an IoU is 0.5. If you want to be strict, you can set a higher threshold, and the higher the threshold is, the better the accuracy is.

The confusion matrix, also known as the error matrix, is a level arrangement used to act for exactness evaluations and is represented by a matrix of n lines and n columns. The specific assessment lists of mistake matrix contain overall accuracy, cartographic accuracy, user accuracy, etc. These exactness lists can show the exactness of image categorization from different aspects.

According to the confusion matrix of the training model, it isn't difficult to draw a conclusion that the system has a total of 6 categories, of which 3 categories have recognition rates above 0.5, and 3 categories have recognition rates below 0.5, namely rock (1), animal (3) and car (6). The identification speeds of these three classes are relatively small. In the subsequent iterations, emphasis will be given to strengthening training of these categories.

In the P diagram of exactness speed, the level coordinate is confidence, and the upright coordinate is accuracy. In general, the accuracy P-graph raises with the increase of trust. After the training of this model, the average accuracy rate of all categories is around 0.901, and most bends indicate a slow rising tendency with the raise of confidence, among which rail has a better accuracy rate. It can be inferred that rail is better than other categories in terms of data set annotation, or it may be because rail has more pictures. Enables the program to recognize specific data more effectively, as shown in Figure 10.

The loss layer is the end point of the CNN. Blob two binary large objects can be entered. Blob contains the predicted result and the actual mark. After carrying out undertakings on the given two inputs, the Loss Function of the existing network, that is, L(θ) is produced, with θ standing for the vector space of the weight in the network. The intent of machine learning is to look for the best L(θ) that is available. Therefore, the weight of θ is the lowest, which can be accomplished by a series of optimization algorithms (such as SGD, etc.). The loss function is obtained through the calculation of forward transfer and is used as the beginning of backward transfer.

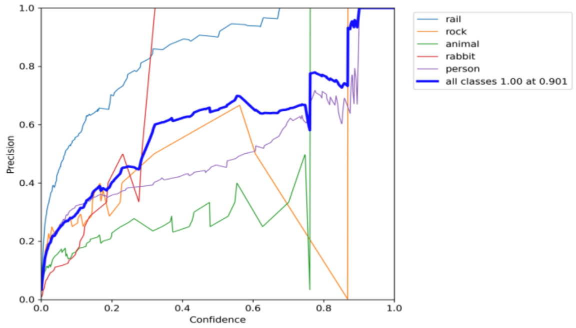


Figure 10 P-Graph

The YOLO series is based on objective scores, class probability scores, and border regression scores to calculate losses. YOLOv5 uses GLOU loss as the boundary box loss, YOLOv5 uses binary cross entropy and partial logarithm loss function to calculate the class probability and target score loss, and we can also use the fl_gamma parameter to activate the focus loss to calculate the loss function.

When the training network is normal, the loss value of the training set decreases and the val_loss value of the test set decreases, which is the most ideal situation. Box_loss refers to the loss function for predicting the position of the box, cls_loss refers to the class loss function, and obj_loss refers to the loss function for whether there are objects in the grid. As can be seen from Figure 11, the box_loss on the extreme left suggests that the training network is functioning correctly. The obj_loss and cls_loss images in the center demonstrate that the loss value of the training set is fluctuating, while the loss value of the test set stabilizes and increases. When the loss value of the test set is stable, the network overfitting can be solved by the following two methods: ① When the data set is not problem, the dropout layer can be added to the "middle depth" of the network or the network depth can be reduced successively; ②When there is a problem with the data set, all the data sets can be shuffled and redistributed. When the loss value of the training set and the loss value of the test set rise at the same time, this is mainly due to the structural design of the network, the setting of instruction hyperparameters, and the definition of the data set.

When the value of the loss function is small, it indicates that the forecasted value is nearer to the true value, producing a better effect. Box_loss is the loss function for the location of the prediction box, and cls_loss is the loss function for the class, and obj_loss refers to the loss function of whether there are objects in the grid. It can be seen from Figure 11 that the position and category of the prediction box in this data set are within a good measurement range, while the training effect of .obj is not good. As can be seen from the graph, the fluctuation is quite drastic.

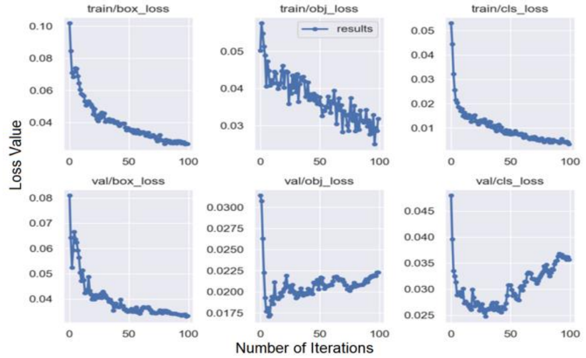


Figure 11 Loss function result curve

5 Conclusion

Based on the demand analysis of the overall system, the feasibility of railway foreign object intrusion monitoring system based on YOLOv5 was determined in this subject. The overall framework and process were designed through the improvement means of computer program engineering, and the program was grown. The railway foreign object intrusion monitoring system has an easy-to-understand interface, which facilitates users to expedite the monitoring process and accomplish the aim of high rate and high effectiveness.

The model recognition rate in the first training is less than 30%. After analysis, the selection of data sets has problems, and the overall data set features are not obvious, which has a great obstacle to training. The model recognition rate of the second training is about 50%. After analysis, it is found that the number of data sets is not enough, and there are fewer types of railway foreign objects, which leads to unsatisfactory overall training effect. After knowing the reason, the whole data set was adjusted, and training file and other parameters have been set. In short, the training model with the recognition rate of 73% was obtained, which was in line with the expected level.

References

[1]Ping H, Ming C, Kexin C, et al.. A combined real-time intelligent fire detection and forecasting approach through cameras based on computer vision method[J]. Process Safety and Environmental Protection,2022,164.

[2]GE Jing. Analysis of the development trend of computer image processing technology [J]. Information Technology Computer Products and circulation.2019(5):45-47.

[3]WANG Quandong, Yang Yue, Luo Yiping, et al.. Review on detection methods of foreign bodies in railway

encroachment [J]. Journal of Railway Science and Engineering,2019,16(12):3152-3159.
[4]Zheng Yuanpan, Li Guangyang, Li Ye. Application of Deep learning in image recognition [J]. Computer Engineering and Applications, 2019,55(12):20-36.
[5]Chen Chao, Qi Feng. Review on the development of Convolutional neural networks and their applications in computer vision [J]. Computer Science, 2019,46(3):63-73.
[6]Chen Rongbao, Zhao Dan, Wang Qianlong. A speed measurement method for vehicles moving ahead based on Image processing [J]. Sensors and Microsystems,2018,37(04):17-19+23.
[7]Xiang Jun, Zhang Jie, Pan Ru Ru et al.. Contour Extraction of Printed Fabric Pattern with Smooth Texture [J]. Journal of Textile Science,2017,38(11): 162-167.
[8]Girshick R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[9]Joseph Redmonl. You Only Look Once: Unified,Real-Time Object Detection.[J]. CoRR,2015,abs/1506.02640.

Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12721

Research Article Open Access

A Review of Lane Detection Based on Deep Learning Methods

Yunzuo ZHANG [1*], Zhiwei TU [1], Fenfen LYU [2]

1. School of Information science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043, China
2. Sports Technical School of Hebei Sports Bureau, Shijiazhuang, Hebei, 050041, China

[*] ***Corresponding Author:*** *Yunzuo ZHANG, E-mail: zhangyunzuo888@sina.com*

Abstract

Lane detection is an important aspect of autonomous driving, aiming to ensure that vehicles accurately understand road structures as well as improve their ability to drive in complex traffic environments. In recent years, lane detection tasks based on deep learning methods have made significant progress in detection accuracy. In this paper, we provide a comprehensive review of deep learning-based lane detection tasks in recent years. First, we introduce the background of the lane detection task, including lane detection, the lane datasets and the factors affecting lane detection. Second, we review the traditional and deep learning methods for lane detection, and analyze their features in detail while classifying the different methods. In the deep learning methods classification section, we explore five main categories, including segmentation-based, object detection, parametric curves, end-to-end, and keypoint-based methods. Then, some typical models are briefly compared and analyzed. Finally, in this paper, based on the comprehensive consideration of current lane detection methods, we put forward the current problems still faced, such as model generalization and computational cost. At the same time, possible future research directions are given for extreme scenarios, model generalization and other issues.

Keywords: *Deep learning; Lane detection; Image segmentation; Object detection; Parametric curves*

1 Introduction

With the continuous development of autonomous driving technology, lane detection, as a key link in application scenarios such as autonomous driving and intelligent parking, plays a crucial role in ensuring that the vehicle accurately understands the road structure and travels in complex traffic environments. After more than a decade of development, lane detection methods [1-2] have evolved through an evolutionary process, from early traditional methods, including image processing [3-4] and feature processing [5], to current deep learning methods [6-9]. Despite the continuous emergence of new detection methods, accuracy and robustness are still the focus of attention in lane detection tasks. With the introduction of deep learning methods, the performance of lane detection has been improved compared to traditional methods, but due to the dataset and other influences, the resulting accuracy is still limited. Although some scholars have proposed corresponding solutions, there are still some limitations.

This paper is devoted to a comprehensive review of lane detection tasks based on deep learning methods in recent years, aiming to provide insights into the latest advances in the field. First, this paper will introduce the background of the lane detection task, including the lane dataset, factors affecting lane detection, as well as traditional lane detection methods and related deep learning methods. By categorizing deep learning methods and analyzing their advantages and disadvantages, it will provide readers with a clear research background.

In the deep learning methods section, this paper will explore segmentation [10], object detection [11], end-to-end, parametric curves and keypoint [12-14] based methods, so as to provide a comprehensive understanding of the characteristics of various lane detection methods. Through a brief comparative analysis of some typical models, this paper will highlight their advantages and limitations in different aspects.

Finally, based on the comprehensive consideration of the current lane detection methods, this paper presents the problems still faced by the current methods, such as computational cost and model generalization, and puts forward some suggestions and solutions to provide the readers with insightful thoughts on the future research direction in this field. Through this review, this paper expects to provide more comprehensive and in-depth references and insights for the research and practical applications in the field of lane detection.

2 Background

As one of the popular tasks in the field of

Received on November 12, 2023; Accepted on December 14, 2023

computer vision, lane detection is widely used in application scenarios such as autonomous driving. With the rapid development of computer vision, lane detection methods have achieved remarkable success in terms of accuracy and robustness. Therefore, an introduction to lane detection and common datasets as well as factors affecting the datasets can help to gain a deeper understanding of the importance and progress of this task.

2.1 Lane detection

In a lane detection task, the vehicle system must be able to accurately capture and represent the shape characteristics of the lane. This usually involves the accurate detection and modeling of various types of lane lines on the road, including solid lines, dashed lines, curves, etc. Accurate lane detection not only contributes to the stability of the vehicle while traveling, but also provides critical decision support such as keeping the vehicle in the correct lane, vehicle turning, etc.

In the process of lane prediction, the shape features of lane lines need to be extracted and represented with high accuracy. Also, for different types of lane lines such as solid and dashed lines, the vehicle system needs to be able to correctly recognize their start and end points as well as curvature information.

Different environmental conditions need to be considered in order to obtain accurate lane shapes, such as lighting changes, bad weather, and nighttime driving. Lane lines may exhibit different visual characteristics in different environments, so the detection system needs to be robust to environmental changes so that it can adapt to a variety of complex real-world situations.

2.2 Common lane datasets

A lane dataset is a set of image data used to train and evaluate lane detection algorithms. These datasets usually contain images of road scenes, which include lane lines, road signs, other traffic signs and environmental information. Such datasets are critical for research and development of lane line detection algorithms because they provide the diversity and complexity of lane lines in the real world. Four common datasets are shown below:

(1) CULane, a dataset provided by the University of Hong Kong, which extracts 133,235 images from 55 hours of Beijing road video data collected by a vehicle-mounted camera and contains image data of traffic scenes under various weather conditions.

(2) TuSimple, 6408 images were extracted by collecting lane images of foreign highways under general weather conditions.

(3) ApolloScape, a dataset provided by Baidu's Apollo autonomous driving team. It provides large-scale image and point cloud data containing images from different scenes such as cities and highways.

(4) BDD100K, a dataset created by Berkeley Deep Learning Lab for autonomous driving scenarios. The dataset contains diverse urban traffic scenarios. The dataset contains a large number of diverse images and videos totaling over 10,000 samples covering different weather conditions, traffic scenarios, and driving environments. The content of the dataset covers various road types such as city roads and highways, and provides rich annotation information including the location and shape of lane lines.

2.3 Factors affecting lane detection

In the field of lane detection, the key factors affecting model performance mainly originate from lane datasets containing information about various road scenes such as city roads, highways, and rural roads [15-16]. While providing training samples, these datasets also carry rich realistic background information, which involves various aspects such as illumination changes [17], object occlusion, and road deformation. In this subsection, we will delve into the factors affecting lane detection in order to fully understand the realistic context of model training. The common influencing factors are as follows:

(1) Influence of lighting changes. Changes in lighting conditions are one of the factors that cannot be ignored in lane detection. Lighting conditions such as strong light, low light, and shadows may lead to changes in the appearance of lane lines, thus affecting the visibility of lane lines.

(2) The effect of object occlusion. In the actual road, vehicles, pedestrians or other obstacles may obscure part of the lane line, making it partially or completely invisible, resulting in the model is difficult to detect the lane completely.

(3) The effect of road deformation. Over time, roads can change the appearance of lane lines due to wear and tear, dirt, rain or snow, increasing the complexity of detection.

(4) The effect of weather changes. Different weather conditions, such as rain, snow or fog, can affect the appearance of the lane or completely cover the lane line, thus affecting the visibility of the lane line.

(5) The impact of complex lanes. Some roads may contain multiple lane lines, intersections or roundabouts, and other complexities, which puts higher demands on lane detection.

3 Traditional Lane Detection Methods

In the development history of lane detection, a series of image processing techniques were used in the early stage, mainly including two aspects of image processing and feature extraction. This section will introduce these two aspects one after another.

3.1 Image processing

Image processing focuses on the alteration of the overall image to adjust the appearance of the image or enhance specific information by using a variety of predefined operations such as filtering, smoothing and edge detection. This approach is robust in manipulating the image as a whole, especially in different environments and scenes. However, it is relatively rigid and not flexible enough to adapt to changes in different tasks, requires manual selection and adjustment of image processing operations, and usually relies on the experience of domain experts. However, with the development of computer vision field, feature extraction methods gradually replace the traditional image processing methods. Among the image processing methods, the main ones are:

(1) Edge detection[18], which detects changes in pixel values in an image in order to determine the edge structure in the image. In lane detection, edge detection helps in identifying the road edges and provides the underlying geometric information.

(2) Color filtering[19], an image processing method that uses color information to emphasize or suppress colors in an image by filtering a specific color channel or setting a color threshold, which helps to emphasize the color of road markings.

(3) Hough transform[20-21], which helps to extract the geometric information of the road by mapping the pixels in the image to the parameter space so as to find the points in the parameter space that share the maximum accumulation to represent the geometric shapes in the image.

(4) Optical flow method[21-22], a technique to obtain motion information by analyzing the motion of pixels in an image between successive frames. In lane detection, the optical flow method can be used to recognize the motion of vehicles and the environment, thus providing information about the location of the lane lines.

(5) Curve fitting[23-24], using a geometric model, is performed by mathematical representation such as polynomials, and optimization techniques such as least squares are used for curve fitting.

3.2 Feature extraction

Feature extraction is concerned with extracting informative local features from an image to capture key information in the image and provide effective input for subsequent tasks. The design of feature extraction is more flexible in that the features used to represent the information in an image can be selected and designed manually. Although manual design of features is required, it is more flexible compared to image processing and does not necessarily require expert level experience. In addition, feature extraction can be computationally expensive depending on the complexity required for the design and computation of the selected features, but is robust in the case of relatively stable local features. Among the feature extraction methods, the main ones are:

(1) Scale-invariant feature transform (SIFT), SIFT is a method for image feature extraction by detecting key points in an image and using local feature descriptors around these key points. In lane detection, SIFT can be used to detect key points in road structure and to describe road texture.

(2) Histogram of oriented gradients (HOG), HOG is a method for extracting image features by analyzing the gradient direction of local regions in an image. In lane detection, HOG can be used to describe the edges and textures in the image to provide strong support for lane detection.

(3) Region of interest (ROI) extraction, Feature extraction is performed by defining and selecting regions of interest in an image and focusing attention on these regions. In lane detection, road regions can be identified by a priori knowledge or image segmentation techniques to reduce computational complexity and increase the reliability of features.

A comparison of image processing and feature extraction features is given in Table 1. The table shows the differences between the two methods by comparing the features of image processing and feature extraction.

Table 1 Features Comparison of Image Processing and Feature Extraction

Characteristics	Image Processing	Feature Extraction
Target	Focus on geometric structure and edge extraction	Capture of local features and textures in images
Robustness	Need to deal with varying environmental conditions	Robust to some local feature stabilizing
Design	Using predefined operations such as filters, edge detectors, etc.	Manual selection and design of feature representations
Cost	Usually faster to compute	Depends on the features design
Flexibility	Not flexible enough to adapt to different tasks	Flexibility to adapt to different tasks
Dependency	Requires human selection and adaptation of image processing operations	Features have to be designed manually
Generalizability	Constrained by a priori knowledge and design	Some generalization capability when local features are stable

4 Deep Learning Based Lane Detection

The emergence of convolutional neural networks [25] marks the shift from traditional methods to deep learning methods [26] in the field of lane detection. Convolutional neural networks bring significant performance improvements to lane detection tasks with their powerful feature extraction capabilities and end-to-end learning. In the face of complex road situations and different environmental changes, CNNs are able to capture and learn the key information in the image more effectively, laying the foundation for accurate and robust lane detection.

4.1 Overview of network model classification

In this subsection, the network models for lane detection based on different methods, including segmentation, object detection, end-to-end, parametric curves and keypoint methods, are explored in depth to reveal in more detail the wide range of applications and advantages of convolutional neural networks in lane detection.

(1) Segmentation methods, through deep learning networks, such as semantic segmentation models, classify images at the pixel level to achieve fine-grained recognition and localization of lane lines. Through convolution and pooling layers, segmentation methods are able to accurately delineate lane regions in an image, providing more precise information for subsequent decision making.

(2) Object detection methods, such as YOLO [27] and Faster R-CNN [28], accurately locate and recognize the lane portion of an image by region detection through convolutional and fully connected layers. object-based detection methods can be categorized as two-stage, single-stage. and anchor-based methods. Choosing the appropriate method can effectively solve the problem of detecting different types of lanes and improve the adaptability and robustness of the model.

(3) The end-to-end methods, by constructing an end-to-end neural network structure, learns the feature representation of lane lines directly from the original image, avoiding the manual feature extraction step in the traditional approach. This approach simplifies the model design process and makes lane detection more intuitive and efficient.

(4) The parametric curves approach, a key step in optimizing the performance of deep learning models. By adjusting the parameters such as the number of layers and nodes of the convolutional neural network, the model is better adapted to different road conditions, thus improving the accuracy of lane detection.

(5) Keypoint methods, which is used to recognize the key feature points of the lane lines in the image, including the start point and the end point. Through convolution and pooling layers, the key point method can effectively capture the important information of lane lines in images.

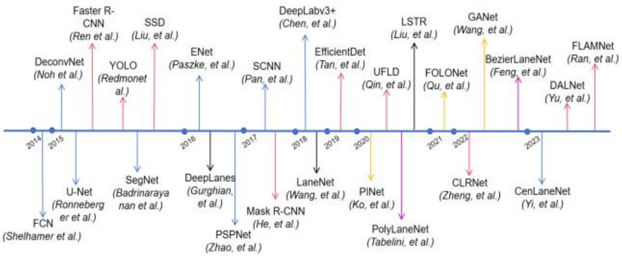


Figure 2 Major methods of deep learning-based network models, including segmentation (in blue), object detection (in red), end-to-end (in black), curve fitting (in pink) and keypoints (in yellow)

In the field of lane lines, many excellent network models have appeared and played an important role. From the traditional convolutional neural networks to the Transformer [29] model that has emerged in recent years, these models provide a variety of options for achieving accurate and efficient lane detection. Figure 2 illustrates typical approaches based on deep learning network models and lists the network models covered in this paper.

4.2 Segmentation methods

Segmentation based methods are based on dividing different pixels into different parts and labeling them according to classification rules. These methods usually include semantic segmentation, instance segmentation and panoramic segmentation.

4.2.1 Semantic segmentation

Semantic segmentation [30-32] aims to assign each pixel in an image to a specific semantic category so that pixels of the same category are labeled as the same region. In lane segmentation, the goal of semantic segmentation is to classify each pixel in an image as belonging to a lane or background. Common semantic segmentation networks include SegNet, U-Net, FCN, and DeconvNet, which achieve pixel-level prediction through convolution and upsampling operations.

In the early days, the use of CNNs greatly simplified the process of extracting image features. In 2014, Jonathan Long et al [33] advanced the field of semantic segmentation by proposing fully convolutional networks. FCN [33] successfully achieved pixel-level semantic segmentation labeling of the entire image through convolution and up-sampling operations. In the following year, a series of network models with encoder-decoder structure emerged to further optimize the semantic segmentation task. Among them, models such as SegNet, DeconvNet, and U-Net became important representatives. SegNet, proposed by Vijay et al [34], utilizes the encoder-decoder structure to regularize the up-sampling and down-sampling operations and improve the accuracy of segmentation. DeconvNet, proposed by Hyeonwoo Noh et al [35], and Olaf U-Net proposed by Ronneberger et al [36] also adopt this structure.

4.2.2 Instance segmentation

The goal of instance segmentation [37] is to split different instances (e.g., different lane lines) in an image and assign unique identifiers to each instance. This method can be used to distinguish between multiple parallel lane lines.

Common lane instance segmentation models include ENet, DeepLabv3+, and LaneNet. ENet was proposed by Adam Paszke et al [38] in 2016, and successfully achieved fast and accurate instance segmentation in resource-constrained scenarios by designing efficient network structures and operations. In the following years, DeepLabv3+ and LaneNet appeared, where DeepLabv3+

is based on an improved version of the DeepLab family, which employs null convolution to expand the sensory field so that it can better capture the contextual information of images. And LaneNet [39], an instance segmentation network focusing on the lane detection task, cleverly combines the ideas of semantic segmentation and instance segmentation, and achieves accurate segmentation of lane lines by assigning each pixel in the image to a lane category and identifying each lane instance. The LaneNet architecture, as depicted in Figure 3, comprises a segmentation branch that creates a binary lane mask, and an embedding branch that generates an N-dimensional embedding for each lane pixel. The embeddings ensure that pixels from the same lane are close together and those from different lanes are far apart. After removing background pixels using the segmentation map, lane embeddings are clustered around their respective centers for identification. Existing models based on instance segmentation still face some challenges on complex scenes in terms of robustness, Yi Sun et al [40] proposed CenLaneNet for robustness. CenLaneNet achieves improved robustness and performance by combining instance segmentation and lane center estimation.

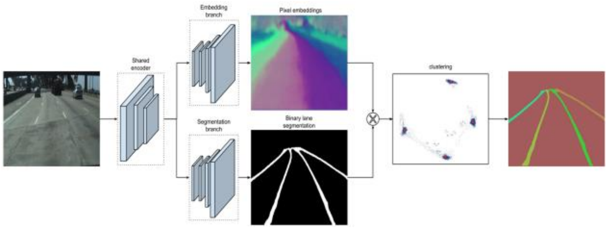


Figure 3 LaneNet architecture framework. The embedding branch is located at the top, while the segmentation branch is located at the bottom. Different lane instances are distinguished using different colors

4.2.3 Panoramic segmentation

Panoramic segmentation [41-42] is the segmentation of an image at a higher level. In lane detection, panoramic segmentation can be used to separate different regions of the entire road (e.g., lanes, sidewalks, lawns, etc.). This helps to understand the road scene more comprehensively. PSPNet [43] is able to capture different scale contextual information and enhance the understanding of the panoramic scene by introducing a pyramid pooling module. The DeepLab family, including DeepLabv3 and DeepLabv3+, allows for detailed panoramic segmentation of the road scene through techniques such as null convolution. UNet employs an encoder -decoder structure that can be used to segment road areas and other environmental elements for panoramic scene understanding.

A comparison of the segmentation-based approaches is shown in Figure 4. the respective goals of interest for semantic, instance and panoramic segmentation are distinguished by color.

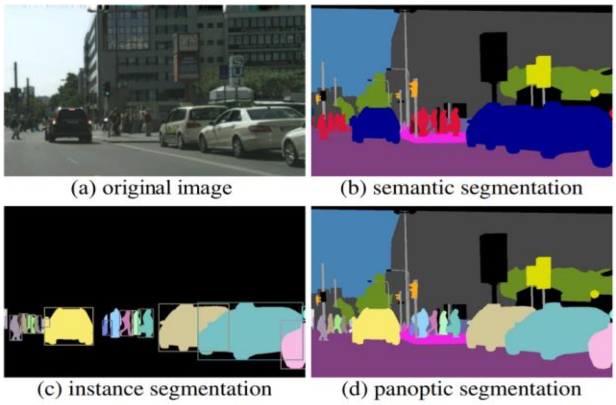


Figure 4 Comparison plot based on segmentation methods. (a) Original image. (b) Region of interest for semantic segmentation. (c) Region of interest for instance segmentation. (d) Region of interest for panoptic segmentation

4.3 Object Detection Methods

4.3.1 Single-stage

Single-stage object detection methods are known for their simplicity and efficiency. These methods usually predict the location and class of the object directly through a single neural network model. YOLO, SSD and EfficientDet are representative algorithms among them.

YOLO [27], as a classical single-stage network, divides the image into grids and performs object detection and localization on each grid while predicting the category of the object. By accomplishing all tasks in a single forward propagation. SSD proposed by W. Liu et al [44] simultaneously predicts objects of different sizes by convolving them at different scales, allowing it to efficiently deal with multi-scale objects. EfficientDet proposed by Mingxing Tang et al [45] focuses on the design of the network in terms of its lightweighting and performance enhancement, which allows it to perform well in the perform well in environments with limited computational resources. It achieves a significant performance improvement on the object detection task by optimizing the network depth and width, as well as adopting a bi-directional feature network structure.

4.3.2 Two-stage

Two-stage object detection methods adopt a staged strategy for object localization and classification. A typical two-stage includes first generating candidate regions and then classifying and regressing the locations of these regions. Mask R-CNN and Faster R-CNN are representatives of two-stage methods.

Mask R-CNN [46],as depicted in Figure 5, achieves instance segmentation based on object detection by introducing additional branches to generate a mask of the object. This enables the model to understand the objects in the image at a finer granularity. Faster R-CNN is a classical two-stage object detection model that generates candidate regions by introducing Region Proposal Network (RPN) and performs classification and location regression in the subsequent stages. It effectively combines the two tasks of candidate generation and object detection, providing a reliable basis for accurate object detection.

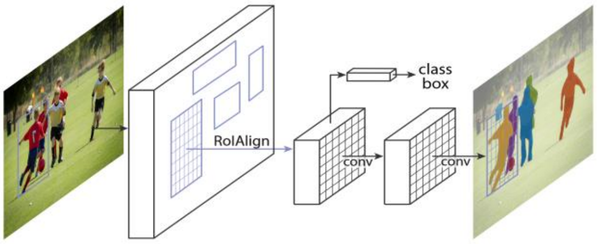


Figure 5 Mask R-CNN framework for instance segmentation

4.3.3 Anchor-based method

In recent years, the application of anchor-based methods has also attracted attention. The main idea of anchor methods is to predefine a set of anchor frames or anchor points through which object detection and localization can be performed. Tu Zheng et al [47] proposed a CLRNet model that first detects lanes with high-level semantic features by combining the lane a priori information, and then performs refinement based on the low-level features [47]. Hao Ran et al [48] proposed FLAMNet, as depicted in Figure 6, with a flexible line anchor mechanism, which enhances local detail extraction and global semantic information modeling by continuously correcting the position of line anchors in order to improve detection performance and computational efficiency. Zichen Yu et al [49] proposed DALNet, a dynamic anchor line-based detection network. This network introduces an innovative dynamic anchor line generator that dynamically generates appropriate anchor lines for each track instance based on the position and shape of the tracks in the input image. These dynamically generated anchor lines not only reflect the positions of the tracks more accurately, but also outperform the predefined anchor lines as positional references. Zequn Qin et al [50] considered lane detection as an anchor-based sequential classification problem, which effectively expands the sensory field and optimizes the complex scene problem by using hybrid anchors to represent the coordinates of the lanes.

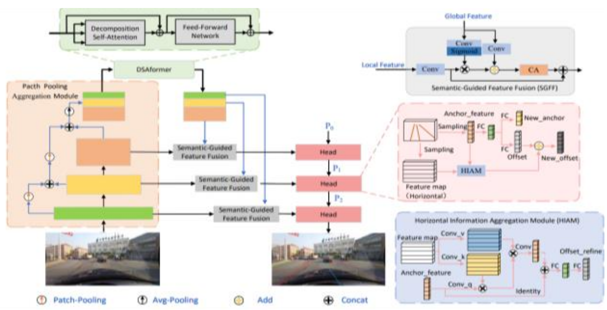


Figure 6 FLAMNet Overview. Features from various layers are pooled and aggregated via the PPAM module. These features then undergo global dependency modeling with the DSAformer, while local and global information fusion is performed using the SGFF module

4.4 End-to-end methods

The end-to-end lane detection method avoids the manual feature extraction step in traditional methods by constructing an end-to-end network model that learns the feature representation of lane lines directly from the original image. This approach enables the model to directly output information such as the location and shape of lane lines.

In recent years, several end-to-end lane detection methods have emerged. Among them, DeepLanes, proposed by Alexandru Gurghian et al [51], employs deep neural networks to reliably estimate the location of lane markers in an end-to-end manner through a classification architecture. On the other hand, Davy Neven et al [52] proposed an instance segmentation method to train an end-to-end detection network containing two branches by transforming the lane detection problem into instance segmentation. In addition, Ruijin Liu et al [53] viewed lane detection as an approximate curve problem, and proposed an end-to-end method that directly outputs the lane shape model parameters, which utilizes the self-attention mechanism of Transformers and is able to learn richer contextual information.

4.5 Parametric curves methods

In lane detection, parametric curve or curve fitting is a common method that describes the shape of lane lines by fitting mathematical curves.

PolyLaneNet, as depicted in Figure 7, proposed by Lucas Tabelini et al [18] uses the idea of polynomial fitting, which describes the curve shape of lane lines by learning polynomial coefficients. This method is more flexible and can adapt to different shapes of lane lines, thus improving the accuracy of detection. LSTR proposed by Ruijin Liu et al [53] combines the ideas of lane line segmentation and tracking, and realizes the parametric curve fitting of lane lines by regression method. The model focuses on modeling the temporal information of lane lines, which in turn improves the adaptability to dynamic scenes. BezierLaneNet proposed by Zhengyang Feng et al [54] uses Bezier curves for fitting lane lines, and represents the shape of the lane lines by learning the control points of the Bezier curves. This method is more flexible in expressing curves and is especially suitable for irregularly shaped lane lines.

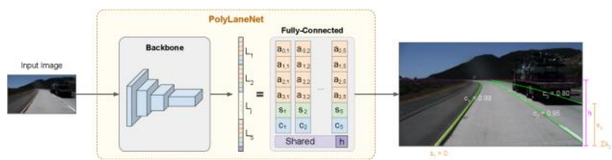


Figure 7 Overview of the PolyLaneNet. From left to right: the model receives as input an image from a forward-looking camera and outputs information about each lane marking in the image

4.6 Keypoint methods

Key point method is a method to achieve lane detection by identifying important feature points of lane lines in an image, such as start and end points. These particular points are critical for understanding the shape and direction of the lane lines, hence the name keypoints. The goal of the keypoint method is to detect and localize these keypoints automatically by using computer vision algorithms or deep learning models to provide detailed information about the location and shape of the lane lines.

In recent years, Yeongmin Ko et al [12] proposed PINet, a deep learning based lane detection method, whose main feature is that it introduces the idea of location awareness and iteration to better capture the location information of lane lines. Through many iterations, PINet gradually optimizes the detection performance for lane lines. In addition, FOLOLane proposed by Z. Qu et al [55] adopts a flow field guided approach to combine the lane detection problem with the flow field task. This unique design allows the model to better understand the dynamic information in the image and improve the accurate detection of lane lines. GANet proposed by Jinsheng Wang et al [56], introduces a novel perspective to lane detection. It directly links each key point to the starting point of its lane line, improving efficiency by eliminating point-by-point extension. This association is achieved by predicting offsets to global lane starting points, enabling parallel processing. Additionally, the Lane Perception Feature Aggregator (LFA) captures local correlations between adjacent key points, enhancing global associations with local information. The GANet architecture is shown in Figure 8.

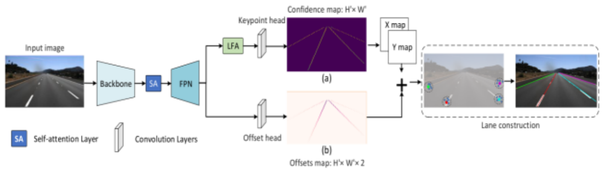


Figure 8 GANet architecture. Confidence map and offset map are combined into group, each representing a lane instance

4.7 Summary of various lane detection algorithms based on deep learning

In this section, we give summarize the approaches, advantages and limitations of representative deep learning based lane detection algorithms. In this paper, we categorize these algorithms into segmentation-based, object detection, end-to-end, parametric curves, and keypoint detection methods. Table 2 demonstrates the representative segmentation methods.

In segmentation-based lane detection methods, this paper divides segmentation methods into three categories according to the goal of processing: semantic segmentation, instance segmentation, and panoramic segmentation. The representative FCN model replaces the fully connected layers with convolutional layers to form a fully convolutional structure, which allows the model to receive input images of arbitrary size and generate dense pixel-level predictions. However, it is worth noting that FCN does not fully consider the relationships between pixels. After the FCN model, the encoder-decoder structure is widely used for semantic segmentation tasks. The SegNet model and the U-Net model are based on the FCN model, which employs an encoder-decoder structure for semantic segmentation tasks. This structure effectively improves the ability to model inter-pixel relationships. DeconvNet employs a convolution-inverse convolution structure, similar to the encoder-decoder structure. For instance segmentation tasks, ENet is a lightweight segmentation network that focuses on high real-time and low latency. However, due to its lightweight structure, for some complex tasks, ENet may perform poorly in terms of adaptability. DeepLabv3+ improves segmentation accuracy by employing null convolution and decoder modules, but requires more computational resources to obtain results. The LaneNet model and CenLaneNet model focus on autonomous driving scenarios, utilizing a multi-branching approach to achieve low Latency. In the panoramic segmentation task, the PSPNet model, based on the FCN model, introduces the pyramid pooling module to establish the connection between local and global features, thus improving the segmentation accuracy. However, the pyramid pooling module increases the computational complexity and memory utilization of the network, resulting in the model requiring more computational resources.

Table 2 Representative segmentation methods

Method	Network	Strengths	Limitations
		Semantic Segmentation	
	Fcn[33]	Using Full Convolutional Structure	Does Not Consider Pixel Relations
	Segnet[34]	Encoder-Decoder Structure	Sensitive To Complex Scenes
	Deconvnet	Up-Sampling Using Deconvolutional Layer	Overfitting
	U-Net[36]	U-Shaped Structure With Encoding And Decoding Processes	Requires More Training
		Instance Segmentation	
Segmentation	Enet[38]	Lightweight Network Structure	Less Generalizable To Complex Scenes
	Deeplabv3+	Introduces Decoder Module	Restricted By Resources
	Lanenet[39]	Focuses On Lane Line Segmentation	Low Scene Generalization
	Cenlanenet [40]	Incorporates Center Line Extraction Idea	Road Detection Needs Improvement
		Panoramic Segmentation	
	Pspnet[43]	Introduces Pyramid Pooling Module	Restricted By Resources
	Deeplabv3+	Panoramic Segmentation Performs Well	Restricted By Resources

In the lane detection method based on object detection, in this paper, we will categorize the lane detection methods based on object detection into single-stage and two-stage approaches, and additionally consider anchor-based methods. Typical single-stage models such as YOLO have fast detection speeds and are relatively weak in small object detection due to the fact that each mesh can only predict one category. SSD employs multi-scale feature maps for detecting objects of different sizes to enhance small object detection, but still performs relatively weakly. EfficientDet uses network scaling to balance the size and performance of the model. Two-stage methods have higher detection accuracy than single-stage methods, but are usually slower and require more computational resources. Typical two-stage methods are Faster R-CNN and Mask R-CNN. Anchor-based method are CLRNet, UFLD and so on. Table 3 demonstrates representative object detection methods.

Table 3 Representative object detection methods

Method	Network	Strengths	Limitations
		Single-stage	
Object detection	YOLO[27]	Fast scenes with high real-time performance	Poor performance on small object inspection
	SSD[44]	Multi-scale feature map detection for variable-size objects	Poor performance on small object inspection
	EfficientDet[45]	Combining efficiency and performance	Poorer performance than some complex models
		Two-stage	
	Mask R-CNN[46]	Introduction of mask segmentation	Complex, high computational requirements
	Faster R-CNN[28]	Classic two-stage	Relatively slow, high hardware requirements
		Anchor-based method	
	CLRNet[47]	Incorporating a priori lane information	May need further adaptability for other object inspection tasks
	FLAMNet[48]	Using line anchors to improve detail and global modeling capabilities	May need further generalization for special scenarios
	DALNet[49]	Allows model to handle different tasks concurrently	Multi-task combination may increase model complexity
	UFLD[50]	Uses lightweight design for model streamlining	Slightly slower to perform in some scenarios

In end-to-end lane detection based approaches, the model maps directly from learning inputs to outputs, simplifying the model tuning process. Representative approaches such as DeepLanes focus on convolutional neural network structure, which reduces manual feature design but is limited by computational resources and dataset size. LaneNet focuses on the lane detection task, which improves the real-time performance of the model, but the training and computational costs are relatively high. LSTR, on the other hand, improves the adaptability to dynamic scenes by introducing a tracking mechanism, but requires more data to train the model and adjust the parameters. Table 4 shows representative end-to-end detection methods.

Table 4 Representative end-to-end detection methods

Method	Network	Strengths	Limitations
End-to-End	DeepLanes[51]	Introduces deep neural network for end-to-end	Variety of training data and scenarios can limit performance
	LaneNet[39]	Focus on lane line detection	Requires more data and computing resources
	LSTR[53]	Directionally output lane shape model parameters	Requires more training data for complex scenarios

Curve fitting methods construct lane shapes by modeling them mathematically. A typical PolyLaneNet model uses polynomials for fitting lanes. Although PolyLaneNet can adapt well to different shapes of lane shapes, the polynomial fitting performance is not flexible enough in complex scenarios such as excessive curvature. Compared with the implementation, BezierLaneNet uses Bessel curves to fit the lanes, which can flexibly adapt to various curve shapes. However, the computation is relatively complex and requires more computational resources. LSTR, on the other hand, can better focus on contextual information to capture lane shapes by introducing a self-attention mechanism, but the model also requires more data to optimize the parameters. Table 5 shows representative parametric curve methods.

Keypoint methods detect lane shape and direction by constructing special points. Typical models include PINet, FOLONet and GANet.

Among them, PINet introduces location awareness and iteration, but there is still room for improvement in its adaptability. It can perform well in ordinary scenes but may be limited when dealing with complex scenes. FOLONet adopts flow field guidance to improve detection accuracy by considering the motion relationship between pixels. Although flow field guidance helps to better understand the dynamic changes in the image, the performance may be low in special scenes. On the other hand, GANet employs generative adversarial networks to improve the robustness and generalization of the model. Although this method performs better in handling complex scenes, the corresponding computational cost is high, which makes it potentially limited in some resource-constrained applications. Table 6 demonstrates representative keypoint methods.

Table 5 Representative parametric curve methods

Method	Network	Strengths	Limitations
Parametric curve	PolyLaneNet[18]	Polynomial fitting of lanes with different shapes	Requires more tuning and optimization for complex scenarios
	LSTR[53]	Focus on global information using self-learning mechanism	Requires more data sets to tune parameters
	BezierLaneNet[54]	Flexibility to fit different curved line shapes	Requires more computing resources when dealing with large data sets

Table 6 Representative key point approach

Method	Network	Strengths	Limitations
Keypoints	PINet[12]	Introduction of location awareness and iteration to incrementally optimize lane line detection performance	Needs to be optimized for complex scenarios
	FOLONet[55]	Introduction of flow field guidance to improve detection accuracy	Lower performance is possible in some special scenarios
	GANet[56]	Introduces generative adversarial network to improve robustness and generalization ability	Higher hardware requirements and higher computational costs

5 Problems and Pspects

Lane detection still faces the impact of datasets and other factors that will directly or indirectly limit the ability of lane detection methods to address complex scenarios, as well as their potential to improve generalization capabilities, real-time performance, and adaptability. The issues of complex scenarios, data diversity, real-time requirements, model generalization, and computational cost are explored in detail below:

(1) In complex scenarios, complex road conditions, multi-lane intersections, road construction and other complex scenarios make the lane detection task more complicated. The lane lines may be curved, bifurcated, widened, etc., which increases the difficulty of the detection algorithms to adapt to irregular shapes. In addition, the existence of a large number of intersections, pedestrian crossings and other scenarios require the lane detection model to maintain high accuracy in these cases, which increases the complexity of the algorithm.

(2) The diversity of data from different geographic locations, road types and traffic environments requires lane detection models to have strong generalization capabilities. Lack of sufficient data training for various complex scenarios can easily lead to model performance degradation in new scenarios, and more samples and more comprehensive datasets are needed to improve the robustness of the model.

(3) Real-time applications such as autonomous driving require high real-time lane detection. Timely and accurate acquisition of lane line information is crucial for decision making and control. Therefore, the model needs to complete the image processing and prediction in a limited time, which puts higher demands on the computational efficiency and speed.

(4) The generalization performance of the model in unknown environments is a key issue in the field of lane detection. Due to changes in actual road and traffic conditions, the model needs to maintain accuracy under different conditions. Fully considering the generalizability of the model so that it can operate stably in various real-world scenarios is an urgent problem to be solved at present.

(5) With the continuous development of deep learning models, some advanced models may have higher requirements on computational resources. The rising computational cost may make it difficult to realize real-time processing in practical applications, especially on embedded devices. Therefore, researchers need to seek more efficient model structures and computational optimization methods while maintaining model performance.

Facing these problems, researchers are constantly seeking innovative solutions. From the algorithmic perspective, some novel deep learning structures and attention mechanisms have been introduced to better handle complex scenarios and improve generalization. In terms of data, more comprehensive and diverse datasets are constructed for training to improve model adaptability. Meanwhile, computational efficiency improvement is also a hotspot in current research, reducing the computational burden through model pruning and lightweight network design.

In future research, more attention needs to be paid to the robustness of the model under extreme conditions, such as extreme weather and light conditions. In addition, further improvement of real-time performance, innovation of data enhancement techniques, and optimization of computational efficiency of deep learning models will be important directions to solve the current problems.

6 Conclusion

This paper provides an exhaustive summary of deep learning-based lane detection algorithms, which is divided into three main parts: First, the background of lane detection is introduced, including lane detection, lane line datasets, and factors affecting lane detection. Second, the traditional methods for lane detection and network models for different deep learning methods are presented, including segmentation-based, object detection, end-to-end, parametric curve, and keypoint methods. In the deep learning methods section, the advantages and limitations of various models are listed in detail. Finally, the complex scenarios, data diversity, real-time requirements, model generalization, and computational costs faced by lane detection are discussed in depth.

Overall, the researchers proposed various innovative solutions, including novel deep learning structures, attention mechanisms, more comprehensive dataset construction, and computational efficiency improvement. Future research directions will focus on addressing model robustness in complex scenarios, improving real-time performance, innovative data augmentation techniques, and optimizing the computational efficiency of deep learning models.

References

[1] Song Ruitao, Li Bin. Surrounding Vehicles'lane change maneuver prediction and detection for intelligent vehicles: a comprehensive review[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23:6046-6062.

[2] Wang Zhangu, Jun Zhan, C. Duan, et al.. A review of vehicle detection techniques for intelligent vehicles[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023,34:3811-3831.

[3] Lin Lianghao. Research on structured road recognition algorithm based on image processing[D]. Harbin University of Science and Technology, 2021.

[4] Zhong Zebin. An image graying method for lane line recognition[J]. Journal of Tongji University (Natural Science Edition), 2019,47:178-182.

[5] Ahmed Hashem, T. Schlechter. Feature-based lane detection algorithms for track following: a comparative study[C]. 2022 International Conference on Mechatronics and Robotics Engineering, 2022.

[6] Yu Fuxing, Wu Yafeng, Suo Yina, et al.. Shallow detail and semantic segmentation combined bilateral network model for lane detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2023,24:8617-8627.

[7] Guo Rui. Research and exploration of lane line detection based on deep learning[C]. International Conference on Image Processing & Computer Vision (IPCV), 2023.

[8] Zhong Haomai. Research and application of lane line detection method based on deep learning[D]. Xi'an University of Technology, 2023.

[9] Tian Sheng, Zhang Jianfeng, Zhang Yutian, et al.. Lane line detection algorithm based on expanded convolutional pyramid network[D]. Journal of Southwest Jiaotong University, 2020,55:386-392.

[10] Li Xiaoxiao, X. Hu, Z. Wang. Advances in instance segmentation research based on deep learning[J]. Computer Engineering and Applications, 2021,57:60-67.

[11] X. Liu, Hu Xiaoguang, Wang Ziqiang, et al.. Progress of anchor-less frame-based object detection method and its application in complex scenes[J]. Journal of Automation, 2023,49:1369-1392.

[12] Yeongmin Ko, Jun Jiwon, Donghwuy Ko, et al.. Key points estimation and point instance segmentation approach for lane detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2021,23:8949-8958.

[13] Lv Chuan. Research on lane line recognition and tracking method based on key point detection[D]. Northwest A&F University, 2021.

[14] Wang Jinsheng, Ma Yinchao, Huang Shaofei, et al.. A keypoint-based global association network for lane detection[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[15] S. Ettinger, Chen Shuyang , Benjamin Caine, Large scale interactive motion forecasting for autonomous driving: the waymo open motion dataset[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021.

[16] Zhang Youcheng, Lu Zongqing, Zhang Xuechen, et al.. Deep learning in lane marking detection: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022,23:5976-5992.

[17] Wang Xi. Lane line detection algorithm in abnormal driving environment[D]. Beijing Jiaotong University, 2022.

[18] L. Tabelini, R. Berriel, T. Paixao. Polylanenet: lane estimation via deep polynomial regression[C]. Proceedings of 2020 25th International Conference on Pattern Recognition, 2020.

[19] Teng Zhu, J. Kim, D. Kang. Real-time lane detection by using multiple cues[C]. International Conference on Control, Automation and Systems, 2010.

[20] S. Kumar, M. Jailia, S. Varshney. An efficient approach for highway lane detection based on the hough transform and kalman filter[J]. Innovative infrastructure solutions, 2022,7(5):290.

[21] G. Taubel, R. Sharma, J. Yang. An experimental study of a lane departure warning system based on the optical flow and hough transform methods [J]. WSEAS Transaction on Systems, 2014,13:105.

[22] W. Kruger, W. Enkelmann, S. Rossle. Real-time estimation and tracking of optical flow vectors for obstacle detection [C].Proceedings of the Intelligent Vehicles' 95. Symposium, 1995.

[23] J. Niu, J. Lu, M. Xu, et al.. Robust lane detection using two-stage feature extraction with curve fitting[J]. Pattern Recognition, 2016,59:225-233.

[24] W. Van Gansbeke, B. Brabandere, D. Neven. End-to-end lane detection through differentiable least-squares fitting[C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[25] A. Krizhevsky, X. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2012,60:84-90.

[26] S. Yalabaka, C. Prasad, P. Sanjana. Lane detection using deep learning techniques[c]. Proceedings of 2022 8th International Conference on Signal Processing and Communication (ICSC), 2022.

[27] J. Redmon, S. Divvala, R. Girshick. You only look once: unified, real-time object detection[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[28] Ren Shaoqin, He Kaiming, R. Girshick, et al.. Faster r-cnn: towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015,39:1137-1149.

[29] A. Vaswani , N. Shazeer, N. Parmar. Attention is all you need[J]. Neural Information Processing Systems, 2017: 5998-6008.

[30] Wang Biyao, Han Yi, Cui Hangbin, et al.. Image-based road semantic segmentation detection method[J]. Journal of Shandong University, 2023,53:37-47.

[31] K. Muhammad, T. Hussain, H. Ullah. Vision-based semantic segmentation in scene understanding for autonomous driving: recent achievements, challenges, and outlooks[J]. IEEE Transactions on Intelligent Transportation Systems, 2022(23):22694-22715.

[32] Tian Xuan, Wang Liang, Ding Qi. A review of deep learning based image semantic segmentation methods[J]. Journal of Software, 2019,30:440-468.

[33] E. Shelhamer, J. Long, T. Darrell. Fully convolutional networks for semantic segmentation[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[34] V. Badrinarayanan, A. Kendall, R. Cipolla. Segnet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017,39:2481-2495.

[35] H. Noh, S. Hong, B. Han. Learning deconvolution network for semantic segmentation[C]. Proceedings of the IEEE international conference on computer vision, 2015.

[36] O. Ronneberger, P. Fischer, T. Brox. U-Net: convolutional networks for biomedical image segmentation[J]. Proceedings of Medical Image Computing and Computer-Assisted Intervention, 2015:234-241.

[37] Yang Shuqin, Ma Yuhao, Fang Mingyu, et al.. Instance segmentation based lane line detection for complex environments[J]. Journal of Zhejiang University (Engineering Edition), 2022,56:809-815.

[38] A. Paszke, A. Chaurasia,S. Kim. Enet: a deep neural network architecture for real-time semantic segmentation[J]. arXiv, 2016,1606.02147.

[39] Wang Ze, Ren Weiqiang. Qiu Qiang. Lanenet: real-time lane detection networks for autonomous driving[J]. arXiv, 2018,1807.01726.

[40] Yi. S, Jun Yu Li, Xu Xin. Adaptive multi-lane detection based on robust instance segmentation for intelligent vehicles[J]. IEEE Transactions on Intelligent Vehicles, 2022,8:888-899.

[41] Xu Pengbin, Qu Anguo, Wang Kunfeng. A review of panoramic segmentation research[J]. Journal of Automation, 2021,47: 549-568.

[42] Wang Hai, Qiu Meng, Cai Yingfeng, et al.. Sparse u-pdp: a unified multi-task framework for panoptic driving perception[J]. IEEE Transactions on Intelligent Transportation Systems, 2023,24:11308-11320.

[43] Zhao Hengshuang, Shi Jianpin, Qi Xiaojuam, et al.. Pyramid scene parsing network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.

[44] W. Liu, D. Anguelov, D. Erhan. Ssd: single shot multibox detector[C]. European Conference on Computer Vision (ECCV), 2016.

[45] Tan Mingxing, Pang Ruoming, Q. Le. Efficientdet: scalable and efficient object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.

[46] He Kaiming, G. Gkioxar, P. Dollár. Mask r-cnn. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)[C]. 2017.

[47] Zheng Tu, Huang Yifei, Liu Yang, et al.. Clrnet: cross layer refinement network for lane detection[C]. Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

[48] Ran Hao, Yin Yunfei, Huang Faliang, et al.. Flamnet: a flexible line anchor mechanism network for lane detection[C]. IEEE Transactions on Intelligent Transportation Systems, 2021.

[49] Yu Zichen, Liu Quanli, Wang Wen, et al.. Dalnet: A rail detection network based on dynamic anchor line[J]. arXiv, 2023,2308.11381.

[50] Qin Zequn, Zhang Pengyi, Li Xi. Ultra fast deep lane detection with hybrid anchor driven ordinal classification[J]. IEEE transactions on pattern analysis and machine intelligence, 2022:1-14.

[51] A. Gurghian, T. Koduri, S. Bailur. Deeplanes: end-to-end lane position estimation using deep neural networks[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016.

[52] D. Neven, B. De, S. Georgoulis. Towards end-to-end lane

detection: an instance segmentation approach[C]. 2018 IEEE intelligent vehicles symposium(IV), 2018.
[53] Liu Tuijin, Yuan Zejian, Liu Tie, et al.. End-to-end lane shape prediction with transformers[C]. Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021.
[54] Feng Zhengyang, Guo Shaohua, Tan Xin, et al.. Rethinking efficient lane detection via curve modeling[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
[55] Z. Qu, Jin Huan, Zhou Yang, et al.. Focus on local: detecting lane marker from bottom up via key point[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[56] Wang Jinsheng, Ma Yinchao, Huang Shaofei. A keypoint-based global association network for lane detection[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

Publisher: Viser Technology Pte. Ltd.

URL: www.viserdata.com

Add.:21 Woodlands Close, #08-18,

Primz Bizhub SINGAPORE (737854)

ISSN 2661-443X