Structured Light Center Extraction Study with Multiple Attention Mechanisms

Hang Sun, Ziwei Zhou*

Abstract—The integration of deep learning with conventional structured light center extraction techniques improves the accuracy of extracting structural gold centers. The method is divided into three steps. The initial step involves calibration, which aims to establish a correlation between image coordinates and world coordinates. The subsequent stage involves identifying the laser fringe area. This study employs a self-designed Multi-Att DeepLabV3+ encoder-decoder neural network architecture to extract the laser fringe region. The self-designed SE-ResSkipNet module is incorporated into the structure as the backbone. The decoder utilizes a parallel alternating dual attention mechanism. The third step involves extracting the center of the laser fringe utilizing the Steger algorithm, which is based on a Hessian matrix. Conducting experimental validation on a laser image dataset that is open-source. The experimental resluts indicate that the network architecture's mIoU, fwIoU, Acc, and Acc class evaluation metrics for complex laser fringe segmentation have shown improvements of 4.22%, 0.67%, 0.33%, and 4.96%, respectively. This algorithm demonstrates superior accuracy compared to other algorithms in laser fringe segmentation, playing a crucial role in the subsequent processes of 3D reconstruction and 3D measurement.

***Index Terms*—Centerline Extraction, Encoder-decoder Network, Attention Mechanism, Steger Algorithm.**

I. Introduction

Significant advancements and progress have been achieved in the fields of medical and industrial design due to the development of 3D scanning and reconstruction technologies. For instance, structured light can be employed in 3D modeling to improve diagnostic procedures. It is utilized for scanning various body parts like teeth, ears, nose, and throat. In industrial design and manufacturing, structured light is extensively applied to measure large objects such as car parts or airplanes through 3D scanning. This method enables the acquisition of more accurate geometric information about the object.

There exist three primary categories of center extraction techniques in structured light imaging: conventional methods [1], improved methods derived from traditional [2] approaches, and deep learning-driven strategies [3]. Various conventional techniques for center extraction include Steger's algorithm, the polar method, edge-based methods, grayscale center of gravity method, curve fitting methods, among others. While these conventional approaches are conceptually simple and relatively straightforward to execute, they are vulnerable to noise interference, resulting in decreased accuracy and stability. The improved approaches, derived from traditional methods, primarily involve the incorporation of advanced techniques and innovative ideas to improve robustness and applicability. In contrast, extraction methods based on deep learning are characterized by their adaptability and generalizability, enabling them to acquire complex representations. Nan et al. [4] proposed an adaptive laser stripe centerline method that utilizes Steger's algorithm. The method efficiently extracts the stripe region by employing the adaptive threshold method. Subsequently, it identifies the edge line of the stripe region using stochastic Hough transform and calculates the width value of the optical stripe region by the normal line. Based on the width value, the optical stripe region is divided, and appropriate parameters are set for Steger's algorithm extraction. The experimental results demonstrate that, in comparison to both the skeleton refinement method and the traditional Steger algorithm, this approach effectively enables the extraction of the centerline of the laser stripe with improved accuracy and stability. Sun et al. [5] introduced a rapid and reliable laser stripe center extraction algorithm that relies on the gray moment algorithm. The method initially establishes a 1D model of light intensity by representing the uniform distribution of laser stripes. Subsequently, it calculates a closed-form solution for the centroid of each laser stripe cross-section based on Legendre's moment conservation law. Finally, the data is smoothed using Reinsch algorithm to preserve the original centroid features and mitigate noise interference. The results demonstrate that the method exhibits strong robustness and accuracy, enabling the effective extraction of the centerline of the laser stripe. The utilization of deep learning in the extraction process can significantly improve the accuracy of the extracted center. Deep learning exhibits robust learning capabilities, enabling it to effectively address challenges encountered by conventional algorithms in noisy and complex environments, showcasing remarkable adaptability. In comparison to conventional algorithms, deep learning models exhibit an advantage in terms of execution efficiency. Following this, Li Yuehua et al. [6] conducted a technique for extracting the center of a light bar using Backpropagation (BP) neural network. The study outlined the fundamental concept of employing BP neural network for light bar center extraction,

Manuscript received September 14, 2023; revised March 1, 2024.

Hang Sun is a postgraduate student of the School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China, e-mail: 1964149327@qq.com.

Ziwei Zhou* is an Associate Professor of the School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (Corresponding author to provide phone: +086-139-4125-5680; e-mail: 381431970@qq.com.)

the process of identifying the optimal center point necessary for network training, and the algorithm for weight adjustments within the network. The study examines the impact of the number of neurons in the hidden layer (m), the number of hidden layers (h), and the training samples on the accuracy of center extraction. This method offers high accuracy and efficiency, making it capable of fulfilling the demands of extracting the center of complex light bars.

In summary, this study proposes a neural network model that integrates encoder-decoder architecture with multi-attention mechanisms to classify pixels within an image, building upon traditional algorithms [7]. The algorithm's task flow for extracting the centerline involves several steps. Firstly, camera calibration is conducted [8]. Secondly, a neural network is utilized to detect the laser streak region [9]. Lastly, Steger's algorithm [10] is employed to extract the pixel point located at the center of the optical streak from the output image. In the subsequent stage, the semantic segmentation model is employed to delineate the laser stripe region. Within the realm of image segmentation, the depth model has been extensively utilized and can serve as a preliminary segmentation model for extracting the light stripe center. However, challenges persist in the realm of laser stripe extraction. These challenges include mitigating the impact of diverse noise sources, addressing issues arising from complex object surface shapes like bending and curving that may cause stripe breakage, overlap, or deformation, consequently compromising the accuracy of extraction outcomes. Additionally, achieving accurate extraction of the stripe center amidst a complex background without interference from other objects remains a significant concern. The semantic segmentation network model is employed to mitigate noise interference and accurately differentiate the pixels within the laser streak region from those in the background region. This approach aims to achieve a highly robust extraction of the laser streak center. In comparison to other network models, the designed model exhibits improved modeling capability, context-aware performance, detail preservation, flexible scalability, and adaptability. It has demonstrated significant achievements in experimental settings. To assess the model's efficacy, this study conducted tests using the open-source laser image dataset from Key Laboratory of Beijing University of Aeronautics and Astronautics (BUAA). The results indicate improvements across all metrics, resulting in increased accuracy.

II. Related Work

A. Camera Calibration

Camera calibration plays a crucial role in the laser streak extraction process by determining both the internal parameters (such as focal length and principal point) and external parameters (such as camera rotation and translation) of the camera. Camera calibration establishes a relationship between image coordinates and world coordinates to ensure accurate 3D reconstruction or measurement. Firstly, it is essential to prepare a specialized calibration board, typically featuring a black and white checkerboard grid or a calibration board designed for a specific pattern. The calibration board features a checkerboard grid or pattern with predetermined dimensions. The calibration plate is subsequently positioned within the camera's field of view to guarantee that it completely occupies the image area. Utilize the camera to capture images from various perspectives and ensure the calibration plate is depicted in diverse poses through position adjustments. Image processing techniques are necessary for extracting the corner positions of each grid point on the calibration plate for every captured image.

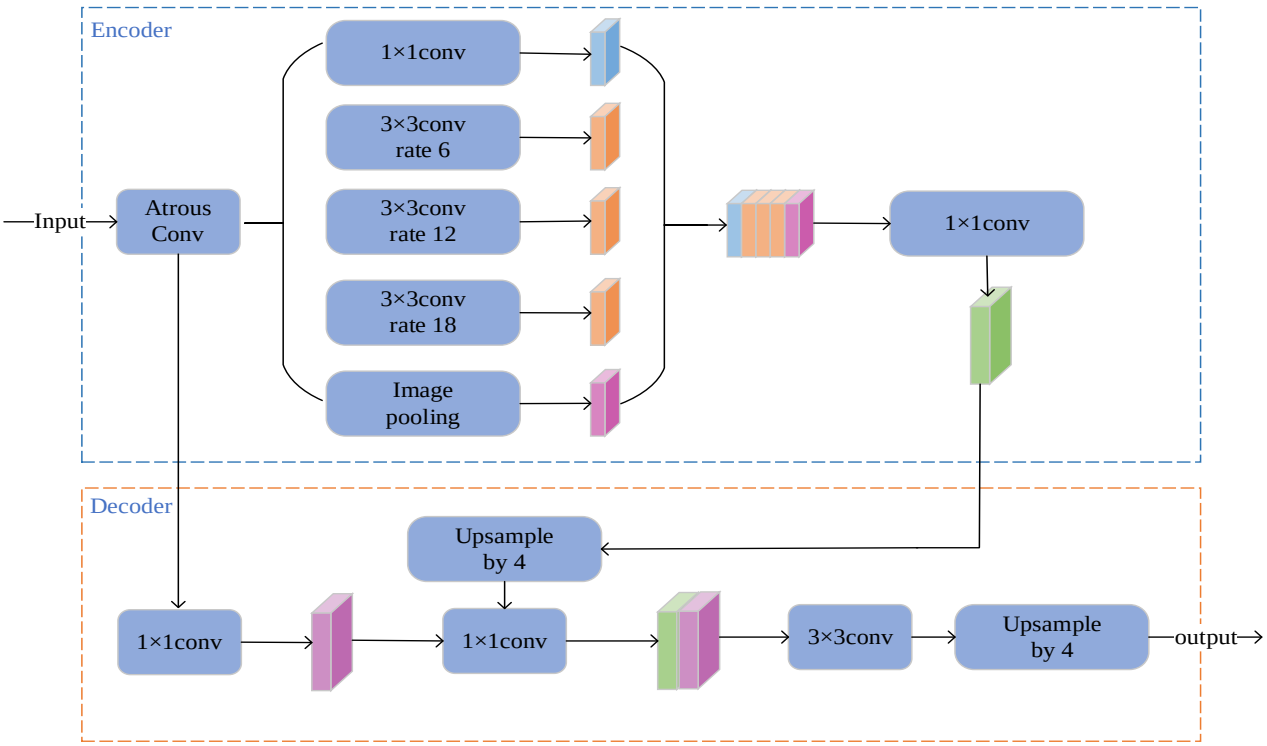


Fig. 1. Structure of DeeplabV3+ neural network model

Establish correspondences between image points and world points by associating the image coordinates of each corner point on the calibration plate with its respective predetermined world coordinates. Utilize the calibration algorithm to estimate the internal and external parameters of the camera based on the known image points. The calibration algorithm is utilized to estimate the internal and external parameters of the camera based on the known image points. Parameter optimization is conducted to improve the accuracy of the calibration.

B. Deeplabv3+ Neural Network Modeling

DeepLabV3+ [11] employs a sturdy backbone network for extracting features from an image, as depicted in Fig. 1. Common choices for this backbone network comprise ResNet, Xception, and other alternatives. The backbone networks have the capability to acquire various levels of feature representation from the input image. Pass the resultant feature maps generated by the backbone networks through an encoder. Typically, the encoder is composed of a series of Dilated Convolution layers.

Cavity convolution expands the receptive field by incorporating voids within the convolution process, facilitating the inclusion of a wider spectrum of contextual details [12]. This improvement improves the overall accuracy of segmentation outcomes. In architectural design, the decoder assumes a crucial role in the process of upscaling the low-resolution feature maps generated by the encoder. Its primary function is to restore these feature maps to align with the resolution of the original image. This process concludes with the production of semantic segmentation results at the pixel level. DeepLabV3+ incorporates Spatial Pyramid Pooling for the purpose of multi-scale feature fusion. This approach entails conducting pooling operations at multiple scales, which enables the acquisition of contextual information across different levels. The feature maps undergo additional processing and are then combined with high-resolution feature maps to produce more accurate semantic segmentation results.

III. Laser Stripe Center Extraction Algorithm Design

A. Camera Calibration Method for Accurate Laser Stripe Center Extraction

The principle of camera imaging is illustrated in Fig. 2. In the 3D world coordinate system $O_W - X_W Y_W Z_W$, a transformation relationship exists between the camera coordinate system $O_C - X_C Y_C Z_C$ and the image coordinate system $o - xy$. Additionally, there is a pixel coordinate system uv used to represent pixel points on an image. Let p represent any point in a real-life scenario, where p is the imaging point, and f denotes the focal length of the camera. By considering a point p in a 3D space and the focal length f of the camera, it is possible to transform the coordinate system and establish a mapping between them. The model is equipped with a normalization plane. Referring to the plane in which the image coordinate system undergoes isometric scaling. Achieving this scaling involves multiplying the coordinates in the image coordinate system by the camera's focal length f . This multiplication by the focal length f facilitates the conversion between the normalization plane and the image coordinate system.

Assuming that point P is represented by the coordinates $X_W = [X_W, Y_W, Z_W]$ in the world coordinate system, $X_C = [X_C, Y_C, Z_C]$ in the camera coordinate system, and $X_u = [X_u, Y_u]$ in the ideal projection plane. The process of transforming from the world coordinate system to the pixel coordinate system in an ideal state is depicted in Equation (1). This transformation achieves the mapping of a 3D scene onto a 2D image.

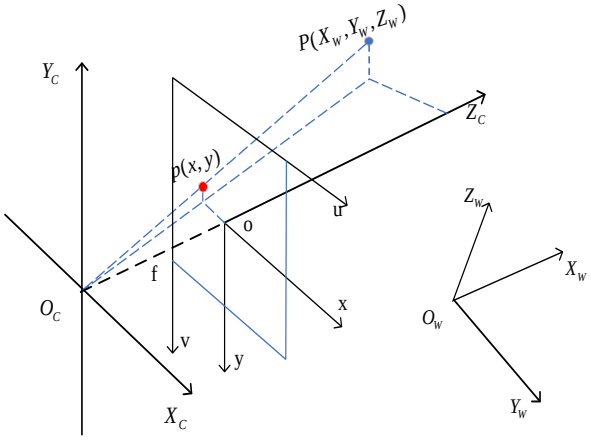


Fig. 2. Schematic of camera imaging principles

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{3\times3} & T_{3\times1} \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{1}$$

where f_x and f_y are the focal lengths in the and directions, respectively; u_0 and v_0 are half size of the calibrated image; $S1$ is the camera's internal reference matrix; and $S2$ is the external reference matrix.

$$S1 = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$S2 = \begin{bmatrix} R_{3\times3} & T_{3\times1} \end{bmatrix} \tag{3}$$

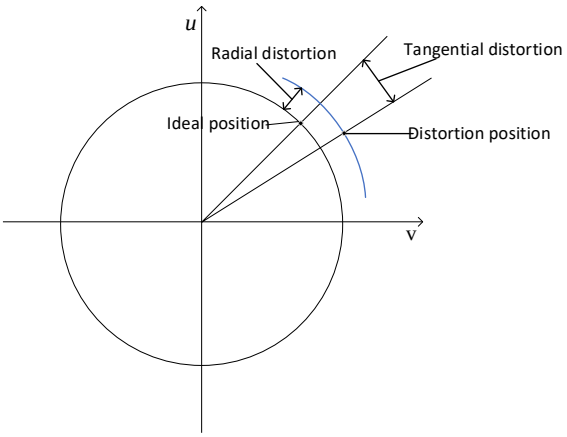


Fig. 3. Schematic of optical distortion in imaging systems

The existence of imperfections in camera lenses and the phenomenon of light refraction lead to the occurrence of aberrations (Fig. 3). These aberrations primarily appear as radial and tangential aberrations. Radial distortion arises from departures from the optimal shape of the lens, leading to the manifestation of curved lines in the image. Tangential aberrations arise due to the misalignment between the lens and the imaging plane, causing straight lines in the captured image to appear distorted or skewed. Correcting these anomalies is crucial for improving the geometric accuracy and visual quality of an image. The aberration parameters, which are essential for the correction process, are acquired via calibration. These parameters delineate the particular aberration traits displayed by the camera lens. As a result, the adjusted image more accurately portrays the shape and size of the object, providing a more realistic and accurate visual representation. Given that aberrations manifest independently and concurrently, radial and tangential aberrations are typically analyzed collectively.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2p_1 xy + p_2(r^2 + 2x^2) \\ 2p_2 xy + p_1(r^2 + 2y^2) \end{bmatrix} \quad (4)$$

where (x', y') are the correct coordinates after removing the distortion; k_1, k_2, k_3 are the radial distortion coefficients; and p_1, p_2 are the tangential distortion coefficients.

The internal and external parameters of the camera, as well as the aberration parameters, can be derived from the aforementioned equations. This enables the computation of the 3D coordinates of the target point, offering essential geometric data required for attaining increased accuracy and usability across various settings.

B. Self-designed Multi-Att DeepLabv3+ Encoder Module

Multi-Att DeepLabV3+ Laser Stripe Detection Neural Network is a convolutional neural network specifically developed for semantic segmentation tasks. The main objective is to accurately extract regions of laser stripes. This network utilizes an encoder-decoder architecture, inspired by DeepLabV3+ model. It optimizes the utilization of the information presented in both the upper and lower text sections, while also guaranteeing the retention of complex details across the divided area.

This study employs ResNet [13] as backbone, which is a deep residual network mitigates the expenses in contrast compared alternative foundational networks. ResNet addresses challenges such as vanishing gradients networks restricted limited capabilities power in deep network by utilizing through connections. This approach sustains maintaining superior capacity for feature extraction features compared conventional convolutional methods. convolutional. ResNet backbone network is utilized for extracting the feature maps. Upon feature map extraction, the utilization of multiscale dilated convolution and Atrous Spatial Pyramid Pooling (ASPP) modules serves as pooling mechanisms to capture elevated-level features. The multiscale dilation convolution technique improves the receptive field by elevating the dilation rate of the convolution kernel. This adjustment allows for the capture of contextual information at various scales. The distinctive features of the laser stripe, which exhibit a significant horizontal scale but a narrow width, highlight the inadequacy of a fixed receptive field in capturing contextual information across different scales. To address this issue, a reduced expansion rate is utilized to effectively capture complex details and small-scale targets. Employing a higher expansion rate is recommended for addressing larger-scale targets and efficiently collecting global contextual information. ASPP module plays a crucial role in this process. The approach involves the utilization of several parallel null convolutional branches, each operating at a different sampling rate, to effectively capture contextual information. Recognizing the limitations of relying exclusively on local information, it becomes evident that the pooling module's effectiveness in addressing diverse scale targets and contextual dependencies is improved through the integration of ASPP module and multiscale dilation convolution. This comprehensive approach ultimately improves the accuracy of the segmentation process.

The proposed model integrates the multi-attention mechanism [14] with long skip connections to improve feature representation. In prior approaches to laser center extraction, distant pixels are influential because of the long-range dependencies involved. Additionally, complex interplays between various channels are essential for specific extraction tasks. Overlooking these connections could result in either loss of information or duplication. Furthermore, the possible loss of essential global information during long-distance information transmission is a matter of concern. To address these challenges effectively, it is essential to integrate attention mechanisms and establish long-range connections within the model. The strategic integration is crucial for effectively managing the complexities presented by various challenges and for enhancing performance and resilience. The integration of attention mechanisms into the model improves the significance attributed to different channels, thereby increasing its ability to focus on these channels. This consequently leads to increased efficiency and improved accuracy in its overall performance.

Specifically, Squeeze-and-Excitation (SE) attention mechanism is employed in the encoder, along with long skip connections. The utilization of SE attention involves aggregating global information onto channels through squeezing and then leveraging the importance of these channels to re-weight the feature map. Additionally, the incorporation of long jump connections enables the bypassing of multiple convolutional layers to directly pass input features to subsequent layers. The seamless integration of the attention module with the convolutional layers facilitates the development of more resilient models, leading to improved feature representation capabilities. Fig. 4 illustrates the architecture of the encoder, comprising the ResNet network along with the parallel-connected dilated convolution and ASPP modules. The multi-scale dilated convolution module includes a 1×1 convolution layer and two 3×3 dilated convolution layers due to the specific structure of the laser bar. Initially, the 1×1 convolution layer is employed to decrease the computational load. Enhancing the feature extraction process can be achieved by incorporating a dilation convolution layer with dilation sizes of 3 and 18, respectively. Large-scale information can be extracted by employing a dilation size of 18, while detailed information can be extracted by using a dilation size of 3.

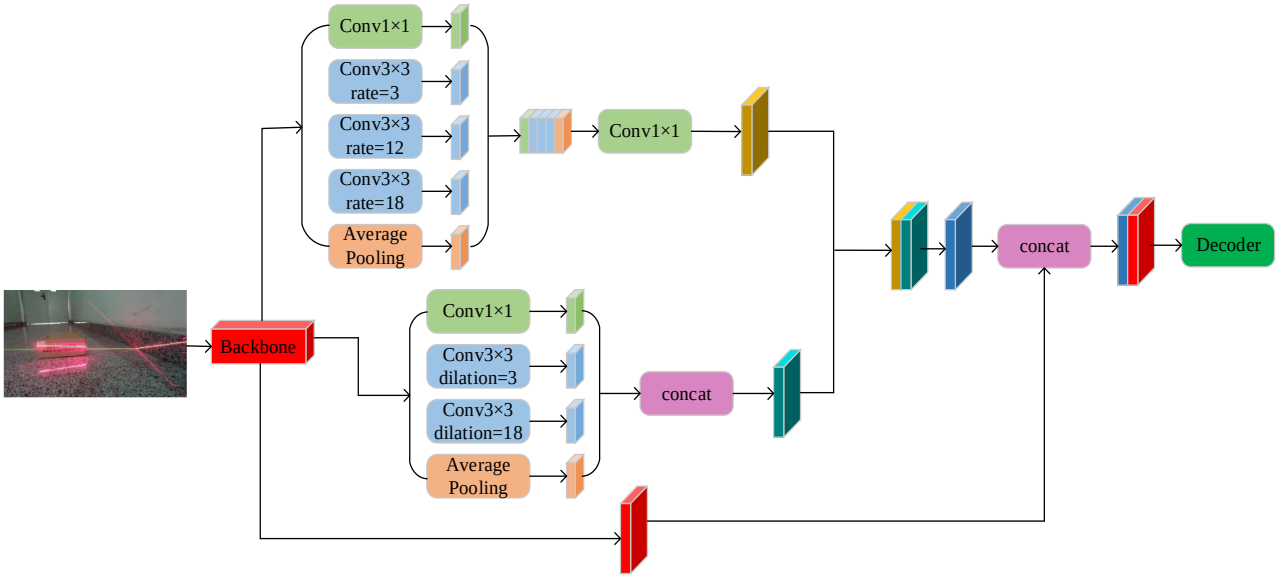


Fig. 4. Structure of Multi-Att DeepLabV3+ encoder

Lastly, incorporate a global average pooling layer. Global pooling enables the aggregation of information from feature channels to extract the overall features of the entire image. Another module in ASPP architecture consists of a 1×1 convolution, a pooling pyramid, and average pooling. The initial branch features a 1×1 standard convolution, a type of convolutional layer that reduces the size of input feature maps to improve computational efficiency and regulate the parameter count. The second through fourth branches consist of ASPP layers with varying expansion rates, designed to improve the receptive field for better capturing contextual information from the laser fringes. This approach aims to improve the performance of fringe extraction. The final layer is the average pooling layer, utilized for integrating the extracted information from various scales. After converting the data into fixed-length feature vectors to maintain spatial invariance, the feature information is integrated using a 1×1 convolutional layer after average pooling. Connecting the two modules in parallel can leverage their individual strengths to extract multi-scale contextual information, integrate and engage with contextual information of varying scales, improve the model's perceptual capabilities, and offer a wealth of global spatial information for laser stripe segmentation [15].

A novel module, SE-ResSkipNet, has been developed through the integration of the SE attention module and skip connections during the extraction of low-level semantic features from the ResNet network, as illustrated in Fig. 5. The architecture comprises a single convolutional layer, followed by two pooling layers, and two attention residual modules. One attention module is characterized by the utilization of two 1×1 convolutional kernels and one 3×3 convolutional kernel, whereas the other attention module employs two 3×3 convolutional kernels. Max pooling and average pooling techniques are utilized to amalgamate the extracted features and diminish the spatial dimensions of the feature maps. The ReLU activation function was selected for its computational efficiency. Subsequently, a fully connected layer is utilized to establish the mapping of the results.

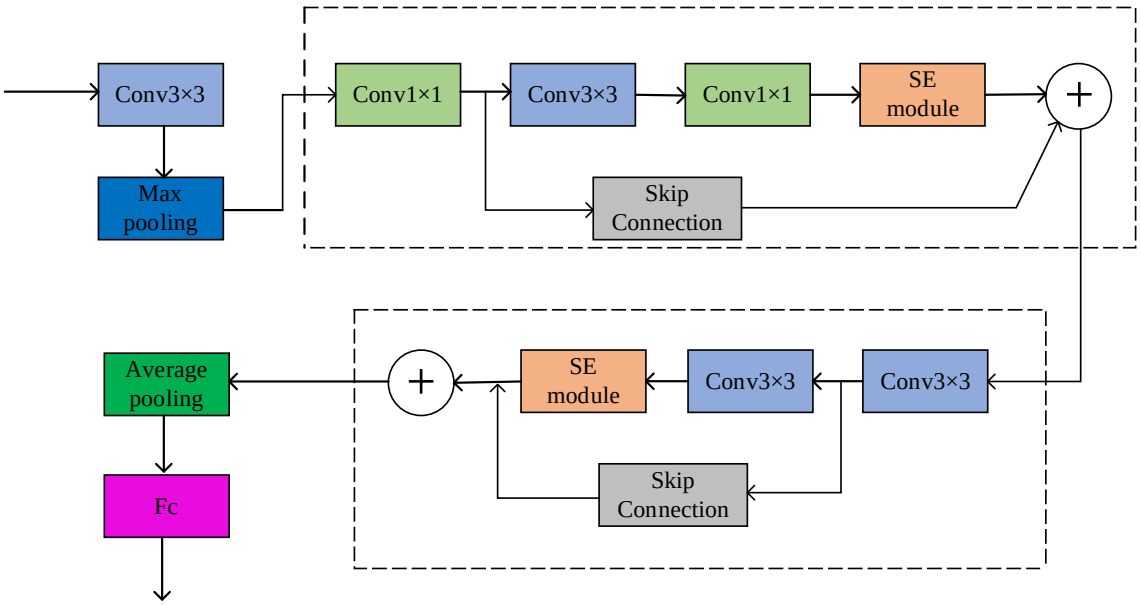


Fig. 5. Structure of SE-ResSkipNet module

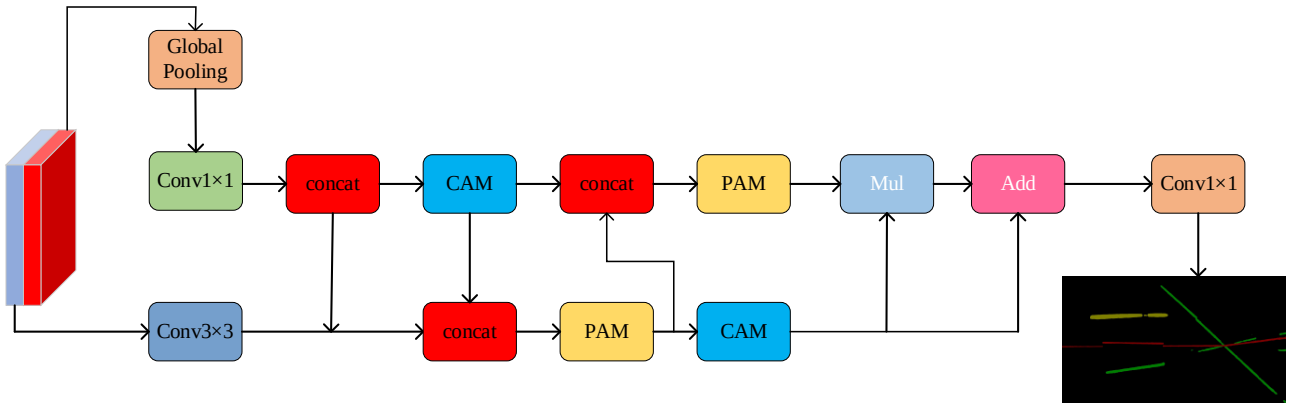


Fig. 6. Structure of Multi-Att DeepLabV3+ decoder

C. *Self-designed Multi-Att DeepLabv3+ Decoder Module*

The decoder module is essential within the system as it is responsible for remapping the abstract features processed by the encoder back to their original spatial dimensions. The design of the decoder concurrently requires the practical retention of fundamental semantic information to maintain semantic consistency between the recovered data and the initial data. Through the decoder function, the system aims to optimize the essential features in the dimensionality reduction process, ensuring that the reconstructed output preserves quality and semantic coherence with the initial input data. The establishment of a solid foundation significantly improves the overall performance and effectiveness of the system.

Fig. 6 depicts the decoder module, comprising the feature fusion module and two convolutional layers. The feature fusion stage integrates Pixel Attention Module (PAM) and Current Amplitude Modulation (CAM) attention mechanisms. PAM attention mechanism modifies the pixel weights by considering the pixel interdependencies within an image. This process improves the representation of correlations among pixels and facilitates the extraction of more significant image features. CAM attention computes channel correlations to derive the weights for each channel and consolidates the weighted features from various channels. This structural approach effectively leverages pixel and channel information, leading to improved feature extraction.

The low-level semantic features extracted by ResNet are combined with the high-level semantic features extracted by the concatenated convolution module based on their respective channel dimensions. Subsequently, the spliced features should be processed utilizing a feature processing module. The convolution layer convolves the input features and incorporates the batch normalization layer to normalize the feature maps. The activation function is applied to the normalized feature maps to activate them. The processed feature maps are subjected to average pooling to reduce spatial dimensions and modify their sizes. The two attention modules are subsequently organized in parallel in an alternating manner, merging the output characteristics of each attention module with those of the two adjacent attention modules. The element-wise multiplication and addition operations interact to combine the processed features with the original features on an element-wise basis, facilitating feature fusion. The element-wise multiplication operation enables the processed features to be multiplied with the original features on an element-by-element basis, thereby highlighting the significance of the processed features. The operation of summing elements one by one involves adding the processed features to the original features on an element-by-element basis, thereby achieving the fusion and combination of features. Through the fusion operation, the information from both the original and processed features are integrated to generate the ultimate fused feature output. The model's performance and accuracy are improved by extracting more sophisticated and expressive feature representations through the fusion operation. The ultimate output is convolved with a 1×1 convolutional layer.

The red region in the resulting figure indicates the location of the light bar in the original image. Consequently, the extraction of the red areas facilitates convenient access to the information pertaining to the light bar in the original image.

D. *Post-processing Algorithms*

The output of the convolutional network is an RGB image consisting of separate red, green, and blue channels [16]. By analyzing individual pixel points in the output image and assigning values according to a specified RGB channel, the exact position of the laser stripes can be identified. Given that the output image aligns with the original image in terms of size and color representation, it effectively filters out noise that is unrelated to the laser fringes. This process simplifies the data processing and isolates the pertinent features, leading to an image that displays the intended laser fringes.

Following the extraction of the specified region, Steger's algorithm, which relies on Hessian matrix [17], is employed to isolate the center of the laser stripe [18]. Determine the normal direction by utilizing the Hessian matrix and apply Taylor polynomials to identify the critical points along the normal. The centroid being sought is the point representing the extreme minimum on the gray distribution curve of the stripe cross-section. This point is characterized by a first-order derivative of 0 and a second-order derivative that is negative. Equation (5) denotes the Hessian matrix corresponding to any pixel point (x, y).

$$H(x,y)=\begin{bmatrix}\frac{\partial^2 g(x,y)}{\partial x^2} & \frac{\partial^2 g(x,y)}{\partial x\partial y}\\ \frac{\partial^2 g(x,y)}{\partial x\partial y} & \frac{\partial^2 g(x,y)}{\partial y^2}\end{bmatrix}\otimes z(x,y)=\begin{bmatrix}r_{xx} & r_{xy}\\ r_{xy} & r_{yy}\end{bmatrix} \quad (5)$$

where $g(x, y)$ is a 2D Gaussian function. When employing Gaussian function, Gaussian filter kernels of diverse sizes and standard deviations are employed to analyze different

regions of the image, thus effectively managing various levels of noise. This approach allows for the fine-tuning and optimization of the filtering effect, leading to improved noise removal and the retention of important image details.

The eigenvalues of the solved Hessian matrix correspond to the poles in the direction normal to the light bar, with the point (x_0, y_0) as the reference, and the sub-pixel coordinates of the center point of the laser bar as (p_x, p_y). The sub-pixel coordinates of the center of the laser bar can be determined by applying equations (6) and (7).

$$(p_x, p_y) = (x_0 + tn_x, y_0 + tn_y) \tag{6}$$

$$t = -\frac{r_x n_x + r_y n_y}{r_{xx} n_x^2 + 2r_{xy} n_x n_y + r_{yy} n_y^2} \tag{7}$$

where r_x and r_y are the first-order partial derivatives of Gaussian function; r_{xx}, r_{xv} and r_{vv} are the second-order partial derivatives of Gaussian function; and n_x and n_y denote the unit vectors in the direction of the stripe normal.

If the offset (tn_x, tn_y) of the center point from the sub-pixel falls within the range of （(-0.5,0.5)，(-0.5,0.5)）for both dimensions, then the point corresponds to the center point of the laser streak image.

The utilization of Steger's algorithm mitigates the error stemming from the uneven grayscale distribution of the optical stripes. The algorithm demonstrates a high level of extraction accuracy and effectively identifies the center of the laser stripes.

IV. Experimental Process

The graphics card model employed in the deep learning segment of this experiment is RTX 3060 Ultra OC server. The operating system environment is the 64-bit Ubuntu 22.04 LTS. The deep learning framework employed is PyTorch. Implement additional programs utilizing Visual Studio 2017 and Matlab. The dataset accessible for this experiment is constrained. The public dataset provided by Key Laboratory of BUAA was chosen for assessing the experimental outcomes, comprising a total of 5,976 images. Divide the data into training and validation sets.

A. Preprocessing of Model Data

Before commencing the training of the model, it is essential to preprocess the dataset. Data improvement techniques such as random flipping, rotating, Gaussian blurring, and adding noise are utilized to augment the images. The images should be labeled according to VOC format, and the evaluation of the proposed model's accuracy should include metrics such as mean Intersection over Union (mIoU), frequency-weighted Intersection over Union (fwIoU), Accuracy (Acc), and Class Accuracy (Acc class). mIoU metric is utilized to quantify the degree of overlap between the predicted segmentation outcomes and the ground truth segmentation results. Given the necessity to prioritize the overall segmentation accuracy, this is achieved through the computation of IoU for individual categories, followed by averaging them. Furthermore, fwIoU serves as an improved segmentation metric, refining mIoU by considering the class distribution within the dataset. Additionally, Acc is utilized to assess the model's categorization accuracy across the complete dataset, while Acc class evaluates the model's categorization accuracy for each specific class.

The equations for mIoU and fwIoU are given as follows:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{8}$$

$$fwIoU = \frac{1}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{9}$$

where k is the number of object categories; p_{ij} is the number of real pixels with number and the number of predicted pixels with number j; $k+1$ is the number of categories; and p_{ii} is the true number.

The equations for Acc and Acc class are expressed as follows:

$$Acc = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{10}$$

$$Acc\ class = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{11}$$

B. Model Training and Evaluation

To mitigate the risk of overfitting. Stochastic Gradient Descent (SGD) was utilized as the optimizer, and Rectified Linear Unit (ReLU) was employed as the activation function for optimization. The training procedure consisted of batches with a size of 8, amounting to a total of 500 iterations. To address the risk of overfitting. In addition to utilizing extensive loss functions like Dice Loss and Cross Entropy Loss Functions, a Dropout Layer was incorporated into the model. The inclusion of a Dropout Layer is essential for reducing the risk of overfitting and improving the generalization capacity of the model, thereby enhancing its performance on various datasets.

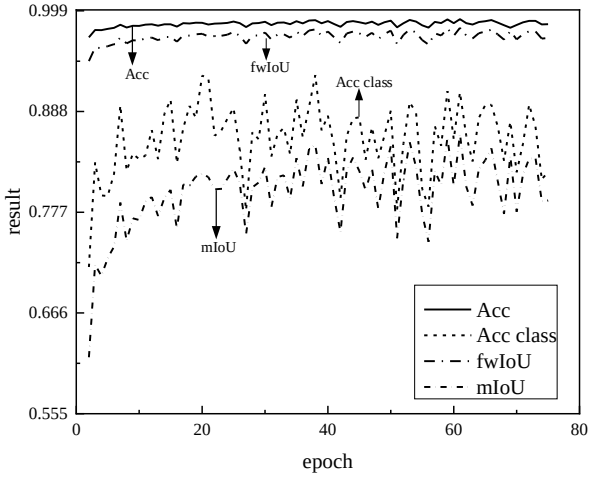
To evaluate the influence of Multi-Att DeepLabV3+ network, ablation experiments were performed on the BUAA dataset. The experiments conducted a comparative analysis of the performance of the network model under study with Latent-Space Dynamic Neural Network (LSDNN) [19] model. The evaluation primarily centered on metrics including mIoU, fwIoU, Acc, and Acc class. After 500 iterations, the results demonstrated significant improvements in all four evaluation metrics. Specifically, there was a 4.22% increase in mIoU, 0.67% in fwIoU, 0.33% in Acc, and a substantial improvement of 4.96% in Acc class. The results highlight the efficacy of Multi-Att DeepLabV3+ network in improving segmentation performance on BUAA dataset when compared to LSDNN model.

TABLE I
EVALUATION METRICS OF ABLATION EXPERIMENTS

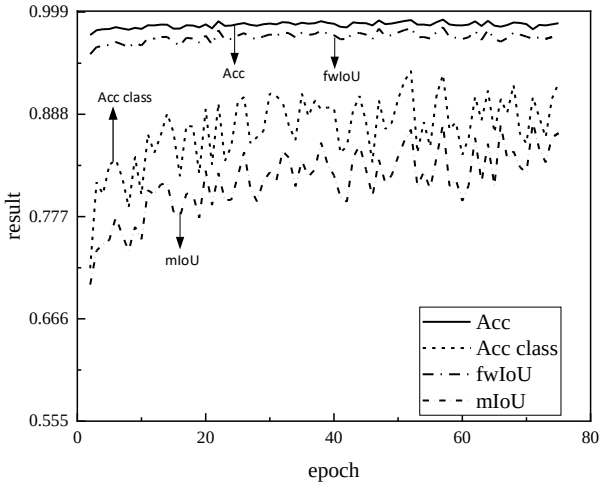
	mIoU	fwIoU	Acc	Acc class
LSDNN	82.76%	97.47%	98.71%	85.97%
LSDNN+SE-ResSkipNet	83.52%	97.87%	98.77%	86.46%
Multi-Att DeepLabv3+	86.98%	98.14%	99.04%	90.93%

Fig. 7 depicts the results of a quantitative assessment conducted through a comparative analysis with the original

model. Seventy-five numerical data points were collected at three-count intervals for this evaluation. Fig. 7(b) displays significant trend characteristics, showing decreased volatility in multiple indicators and an overall upward trend, indicating outstanding performance. The results demonstrate a significant improvement in the performance of the proposed model. These observations provide additional support for the conclusions regarding the improved performance of the new model and lay a robust groundwork for this research.



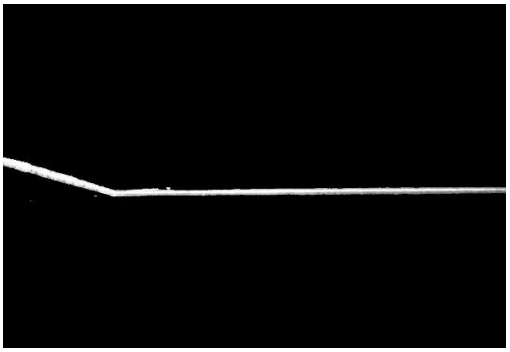
(a) LSDNN



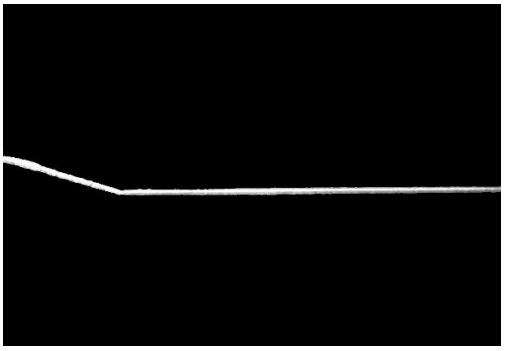
(b) Multi-Att DeepLabv3+

Fig. 7. Quantitative comparison result

Fig. 8 illustrates a comparison between the output image of the original network model and the obtained result. The laser stripes exhibit a more refined appearance in the output of this study.



(a) LSDNN



(b) Multi-Att DeepLabv3+

Fig. 8. Laser strip segmentation comparison

To ensure a comprehensive comparison in the proposed experiments, this study meticulously chose developed models for semantic segmentation, including U-Net [20], FCN [21], and SegNet [22]. To uphold the credibility and impartiality of the experimental data, the consistency in the hardware and software environment was ensured for both training and testing procedures. The experiments utilized consistent learning strategies and carried out an equal number of iterations. The table below illustrates a comparison of mIoU results obtained after an equivalent number of iterations, showcasing the improved accuracy of the proposed approach.

TABLE II
SEGMENTATION ACCURACY ACROSS DIFFERENT METHODS

	mIoU
U-Net	83.24%
FCN	84.36%
SegNet	82.43%
Multi-Att DeepLabv3+	86.98%

In the proposed post-processing algorithm, this study conducted a comparison between the Steger method and alternative techniques such as the directional template and grayscale center of gravity methods. The results, as illustrated in Fig. 9, indicate the superior performance of the Steger method in accuracy and the continuity of centerline extraction.

(a) Direction template (b) Steger Method (c) Grayscale center of gravity

Fig. 9. Laser strip center extraction comparison

To assess the computational complexity, a comparison was carried out of the extraction efficiency across the three methods. The table presented below illustrates the average time taken by each method for extracting the center of optical strip images.

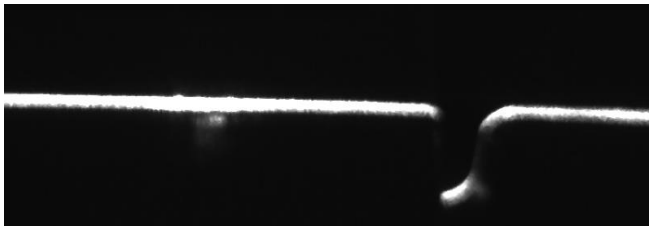
TABLE III
TIME EFFICIENCY COMPARISON FOR DIFFERENT EXTRACTION METHODS

Method	Direction Template	Steger Method	grayscale center of gravity
Time	0.347	0.332	0.182

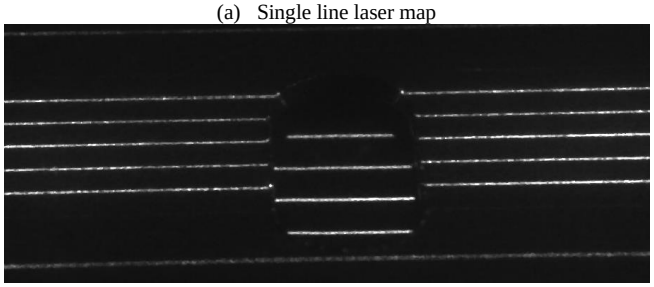
While the algorithm employed in this study may exhibit

reduced running speed, it showcases improved performance in terms of extraction accuracy. The algorithm utilized in this context attains a favorable balance between speed and accuracy, thereby establishing a competitive edge. The prioritization of extraction accuracy over speed highlights its appropriateness for applications that value accuracy and high-quality outcomes. Therefore, this study employs a combination of deep learning algorithm and Steger algorithm to improve the accuracy and resilience of laser centerline extraction. This approach facilitates better feature information capture and enables seamless adaptation to diverse scenarios and demands.

To ascertain the reliability of the proposed method, a comprehensive experimental analysis of the improved algorithm is conducted across various dimensions. The laser profile and multi-line laser image captured are presented below.



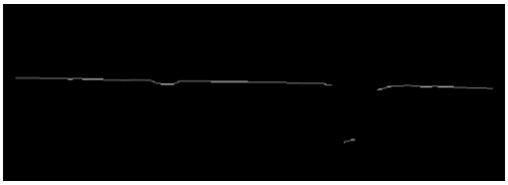
(a) Single line laser map



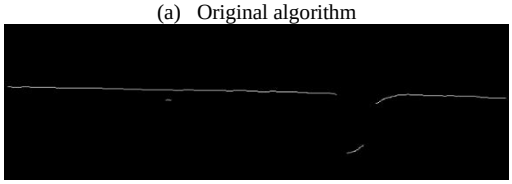
(b) Multiline laser map

Fig. 10. Laser strip image

The study compares the impact of the original algorithm and the improved algorithm by extracting the central line from the single-line laser image. The optimized algorithm centerline demonstrates a superior alignment with the laser stripe image in the original image. This improved effect is evident not only in the accuracy of the central line fitting but also in the concentration of the outcomes. The improved algorithm improves the focus of the center line of the laser fringe, thereby facilitating a more accurate analysis of the deformation characteristics of the object surface. The results are depicted in Fig. 11 and 12.



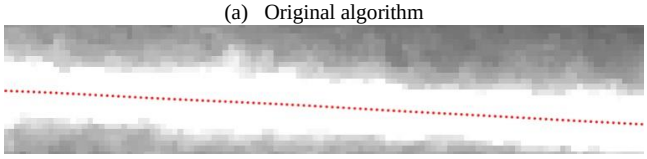
(a) Original algorithm



(b) Improved algorithm

Fig. 11. Extraction effect comparison

(a) Original algorithm



(b) Improved algorithm

Fig. 12. Centerline extraction effect comparison

The comparison between the running time of the original algorithm and the improved algorithm introduced in this study is conducted due to the elevated computational complexity of the algorithm. The objective of this study is to minimize the computational burden, reduce the runtime, and improve real-time performance through algorithm optimization. Table IV presents the average time taken by the two algorithms for the multiple extraction of five laser fringe images.

TABLE IV
COMPARATIVE ANALYSIS OF CENTERLINE EXTRACTION TIMES

	1	2	3	4	5
Original method	0.681	0.836	0.815	0.782	0.764
Improved method	0.332	0.413	0.396	0.383	0.316

The comparison results above indicate that the algorithm demonstrates higher accuracy in extracting the center of the laser line and significantly reduces the processing time. However, the extent of time reduction is dependent on the number of pixels involved. The algorithm can offer more dependable and accurate data for future practical applications based on this outcome.

V. Conclusion

To improve the efficacy of laser centerline extraction, this study employs Multi-Att DeepLabV3+ model for detecting optical strip images. This approach aims to mitigate the adverse effects of diverse interference information on the accuracy of optical strip extraction. The Steger algorithm is subsequently employed to extract the center of the laser stripe. The incorporation of deep learning models improves the flexibility of this approach in capturing complexity in laser images, thereby enabling more accurate extraction of laser centerlines. This advancement offers robust support for the field of laser measurement. Multi-Att DeepLabV3+ model incorporates the innovative SE-ResSkipNet convolution module, which has been applied and validated using the publicly available laser fringe dataset. The experimental results demonstrate that the suggested approach has improved the accuracy of laser fringe segmentation. The evluation metrics of mIoU, fwIoU, Acc, and Acc class demonstrated an increase of 4.22%, 0.67%, 0.33%, and 4.96%, respectively. The experimental results demonstrate the efficacy of the self-designed model in accurately extracting laser line centers, showcasing clear advantages over alternative methodologies. The method employed in this study for extracting the entire centerline is deemed to be more accurate and resilient, thereby offering more accurate data for future applications.

REFERENCES

[1] Y. Wang, Q. Chen, Q. Zhu, L. Liu, C. Li, and D. Zheng, "A survey of mobile laser scanning applications and key techniques over urban areas," *Remote Sensing*, vol. 11, no. 13, pp. 1540, 2019.

[2] Y. Zhou, X. Q. Men, and D. B. Jiang, "Structured light streak center extraction method under complex interference," *Chinese Journal of Lasers*, vol. 47, no. 12, pp. 1204004, 2020.

[3] S. K. Zhang, Q. X. Wu, and Z. Y. Lin, "Detection and segmentation of structured light streaks in weld images," *Acta Optica Sinica*, vol. 41, no. 5, pp. 0515002, 2021.

[4] F. Nan, D. H. Li, Q. Gao, and X. Yu, "Adaptive optical stripe center extraction with improved Steger algorithm," *Laser Journal*, vol. 39, no. 1, pp. 85-88, 2018.

[5] Q. Sun, R. Liu, and F. Yu, "A laser stripe center extraction method based on Legendre moment," *Chinese Optics*, vol. 127, no. 2, pp. 912-915, 2016.

[6] Y. H. Li, P. Liu, J. B. Zhou, Y. Z. Ren, and J. Y. Jin, "Structured light bar center extraction based on BP neural network," *Acta Optica Sinica*, vol. 39, no. 12, pp. 1212005, 2019.

[7] J. H. Mei, and L. J. Lai, "Research on optical strip center extraction algorithm of structured Light measurement System," *Electronic Measurement Technology*, vol. 41, no. 21, pp. 8-13, 2018.

[8] G. H. Wang, and K. M. Qian, "A review of calibration methods for line array cameras," *Acta Optica Sinica*, vol. 40, no. 1, pp. 0111011, 2020.

[9] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L. C. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12475-12485, 2020.

[10] J. Liu, and L. H. Liu, "Laser fringe center extraction based on Hessian matrix and region growth," *Laser & Optoelectronics Progress*, vol. 56, no. 2, pp. 021203, 2019.

[11] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroof, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818, 2018.

[12] B. Ma, and Y. Chen, "Attentive Enhanced Convolutional Neural Network for Point Cloud Analysis," *IAENG International Journal of Computer Science*, vol. 50, no.2, pp417-421, 2023.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[14] C. J. Zhang, L. Zhu, and L. Yu, "Review of attention mechanism in convolutional neural networks," *Computer Engineering and Applications*, vol. 57, no. 20, pp. 64-72, 2021.

[15] F. Nie, and J. Li, "Image Threshold Segmentation with Jensen-Shannon Divergence and Its Application," *IAENG International Journal of Computer Science*, vol. 49, no.1, pp200-206, 2022.

[16] P. Chmelar, and M. Dobrovolny, "The laser line detection for autonomous mapping based on color segmentation," *International Journal of Mechanical and Materials Engineering*, vol. 7, no. 12, pp. 1654-1658, 2013.

[17] Q. Bao, and F. Li, "Light fringe center Extraction method based on Hessian Matrix combined with gradient variance," *Computer and digital engineering*, vol. 48, no. 8, pp. 2018-2023, 2020.

[18] W. M. Li, G. Peng, X. Y. Gao, and C. Ding, "A fast extraction algorithm for the center of line laser light bar," *Chinese Journal of Lasers*, vol. 47, no. 3, pp. 304002, 2020.

[19] C. Zhao, J. Yang, F. Zhou, J. Sun, X. Li, and W. Xie, "A robust laser stripe extraction method for structured-light vision sensing," *Sensors*, vol. 20, no. 16, pp. 4544, 2020.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Processing of the MICCAI conference on Medical Image Computing and Computer*, pp. 234-241, 2015.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.

[22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.

HANG SUN obtained B. Eng. degree in software engineering from University of Science and Technology Liaoning, Anshan, China, in 2021. She is currently pursuing an M. Eng. degree at School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, China. Her research focuses on computer vision and deep learning.

ZIWEI ZHOU obtained B.Sc. and M.Sc. degrees in computer science from University of Science and Technology Liaoning, China, in 1997 and 2005, respectively, and Ph.D. degree in control science from the Harbin Institute of Technology, China, in 2013. He holds the position of Associate Professor and serves as a master's Supervisor at the School of Electronic and Information Engineering, University of Science and Technology Liaoning. His primary areas of research focus encompass deep learning, image processing, and robot control systems.