AI Method for LAMOST Fiber Detection Based on Front Illumination (Postprint)

Authors: Zhangze Chen, Yihan Song, Ali Luo, Guanru Lv and Haotong Zhang

Date: 2025-06-13T16:50:50+00:00

Abstract

The double revolving fiber positioning technology employed in the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) represents one of the most successful advancements in large-scale multi-objective spectroscopy. The precision of fiber positioning is crucial, as it directly impacts the observational efficiency of LAMOST. A critical component of the fiber positioning system is the closed-loop control system, which traditionally utilizes the light spot generated at fiber end. However, this study introduces a novel approach based on front-illuminated LAMOST focal plane image measurements. Unlike back-illumination, front-illumination does not necessitate internal lighting in the spectrograph, thus reducing light pollution and eliminating the need for additional photography. This method employs an artificial intelligence model to analyze images captured at the focal plane unit (FPU), using the image of the white ceramic head on the FPU as the data set for training, the model is capable of accurately measuring the fiber positions solely through front-illumination. Preliminary trials indicate that the measurement accuracy achieved using the front-illumination method is approximately 0. 13. This level of precision meets the stringent fiber positioning accuracy requirement of LAMOST, set at 0. 2. Furthermore, this novel approach demonstrates compatibility with LAMOST's existing closed-loop fiber control system, offering potential for seamless integration and enhanced operational efficiency.

Full Text

Preamble

Research in Astronomy and Astrophysics, 25:065004 (13pp), 2025 June

https://doi.org/10.1088/1674-4527/adce4d
CSTR: 32081.14.RAA.adce4d

AI Method for LAMOST Fiber Detection Based on Front Illumination

Zhangze Chen[1,2,3], Yihan Song[1,3], Ali Luo[1,3], Guanru Lv[1,3], and Haotong Zhang[1,3]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; yhsong@nao.cas.cn, lal@nao.cas.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Received 2024 December 26; revised 2025 March 4; accepted 2025 April 8; published 2025 May 23

Abstract

The double revolving fiber positioning technology employed in the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) represents one of the most successful advancements in large-scale multi-object spectroscopy. The precision of fiber positioning is crucial, as it directly impacts the observational efficiency of LAMOST. A critical component of the fiber positioning system is the closed-loop control system, which traditionally utilizes the light spot generated at the fiber end. However, this study introduces a novel approach based on front-illuminated LAMOST focal plane image measurements. Unlike back-illumination, front-illumination does not necessitate internal lighting in the spectrograph, thus reducing light pollution and eliminating the need for additional photography. This method employs an artificial intelligence model to analyze images captured at the focal plane unit (FPU), using the image of the white ceramic head on the FPU as the dataset for training. The model is capable of accurately measuring fiber positions solely through front-illumination. Preliminary trials indicate that the measurement accuracy achieved using the front-illumination method is approximately 0.13 arcseconds. This level of precision meets the stringent fiber positioning accuracy requirement of LAMOST, set at 0.2 arcseconds. Furthermore, this novel approach demonstrates compatibility with LAMOST's existing closed-loop fiber control system, offering potential for seamless integration and enhanced operational efficiency.

Key words: instrumentation: detectors – techniques: image processing – methods: data analysis

1. Introduction

Multi-object fiber surveys have significantly enhanced the efficiency of astronomical observations by allowing the simultaneous observation of multiple celestial objects. This advancement is achieved through the deployment of multiple movable fiber positioning units. Automation technology is now integral to large-scale multifiber spectral surveys, providing precise control and detection of the fibers.

A notable development in this field is the parallel controllable fiber positioning system (Xing et al. 1998), which utilizes double revolving fiber positioning units (FPUs).

This innovative system was first implemented in the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012). LAMOST can position 4000 fibers with an impressive accuracy of 0.4 arcseconds within only 10 minutes, marking a significant milestone in the automation and precision of astronomical surveys. Subsequent surveys, including the Dark Energy Spectroscopic Instrument (DESI; Schubnell et al. 2016), the Multi-Object Optical and Near-Infrared Spectrograph (MOONS; Montgomery et al. 2016), and the Prime Focus Spectrograph (PFS; Fisher et al. 2014) for the Subaru Telescope, have all adopted fiber positioning systems analogous to the one utilized by LAMOST.

The new surveys usually use a visual closed-loop system to detect the fiber position more precisely, which is always equipped with a monitor camera, e.g., PFS (Wang et al. 2016), MOONS (Drass et al. 2016) and DESI (Baltay et al. 2019). However, LAMOST currently employs a semi-open loop system, which does not provide a method to confirm whether the fiber has moved to the correct position during observations. To mitigate this, LAMOST calibrates the coordinates of the FPUs at the beginning of each observation season, ensuring precise zero positions and accurate movement trajectories. Additionally, the system requires some time to return the FPUs to their pre-calibrated home positions before moving to new positions. Despite these efforts, occasional mechanical failures can lead to collisions between adjacent FPUs, potentially impacting fiber positioning accuracy and risking damage to the FPUs themselves. The calibration and repositioning process incurs additional time and costs due to the lack of a real-time system for detecting the exact fiber location. LAMOST is now actively pursuing the implementation of a closed-loop real-time system to enhance observation precision, and already has a research program based on back-illumination (Zhou et al. 2018).

The closed-loop system primarily comprises two categories of methods: the back-illuminated method and the front-illuminated method. The back-illuminated method is a well-established technique and has been widely implemented in several multi-object fiber surveys, such as DESI (Baltay et al. 2019) and PFS (Wang et al. 2016). This method operates by emitting light from the end of the spectrometer, which then travels through the optical fiber and exits at the focal plane. A high-precision camera captures the emitted light to calculate the position of the optical fiber with high accuracy. However, this method requires a complex back-illumination system. For instance, when LAMOST attempts to construct its back-illuminated system, it must include a slit of LED bulbs that matches the fiber slit and a device to move the LED slit in and out of the light path. This movement during closed-loop control introduces potential instability due to an increased risk of mechanical errors. Furthermore, the presence of light inside the spectrograph poses a risk of light pollution, with the potential for overwhelming photons to cause CCD saturation and residual electric charge erase. Additionally, the mechanical movement of the device increases overhead time during observations.

The front-illumination technique, as implemented in the MOONS project (Drass et al. 2016) and methodologically explored in LAMOST (Zhou et al. 2022), provides a direct approach to fiber positioning through the capture of illuminated focal plane images. This allows for the complete acquisition of the FPUs' mechanical structure, which is then used to calculate the fiber positions. In addition to determining the fiber positions, it is essential to ascertain the posture of the FPUs to prevent mechanical collisions during LAMOST observations (Zhou et al. 2021). Consequently, for the back-illuminated method, two captures are required: one back-illuminated shot for fiber positions and one front-illuminated shot for FPU posture, resulting in additional time consumption and reduced observation time. In contrast, the front-illuminated method is more time-efficient. However, achieving the necessary accuracy with this method is challenging. To precisely determine the fiber positions, the method demands either a complex hardware mechanical structure, such as dedicated metrology targets atop the FPU (Drass et al. 2016), or a sophisticated algorithm for analyzing the FPUs' mechanical structure (Zhou et al. 2022).

For LAMOST, the lack of a dedicated front-illuminated detection design necessitates the development of an advanced algorithm for precise fiber position detection. This algorithm is designed to accurately locate fiber positions by detecting the pinhole of the fiber (as shown in Figure 1), achieving the necessary precision despite numerous environmental factors impacting actual observations. Accurately identifying the pinhole of each fiber within a complex scene containing thousands of FPUs presents a significant challenge. Our goal is to attain pinpoint accuracy within a margin of error of 0.2 arcseconds, equivalent to approximately 0.17 pixels. Additionally, compensating for deviations due to the camera's angle relative to the focal plane and managing the varying orientations of FPUs during operation further complicate the task. Our research indicates that front-illuminated images contain sufficient information to utilize artificial intelligence (AI) technology for fiber detection. This study focuses on employing AI methods to achieve front-illuminated closed-loop control of the telescope through modifications to the software algorithm alone, thereby exploring new methodologies for future multi-object fiber surveys.

The structure of this paper is as follows: Section 2 outlines the methodologies employed and provides detailed descriptions of each. Section 3 presents the results obtained from our study, accompanied by a comparative analysis. Finally, Section 4 discusses the limitations of the study, while Section 5 provides the conclusions.

2. Method

In this study, we have developed an AI-based methodology capable of accurately determining the real-time positions of optical fibers using only front-illuminated images. This method adheres to the precision standards required by the LAMOST optical fiber closed-loop control system. The overall workflow of our approach is illustrated in Figure 2. This section will provide a detailed exposition of our method.

2.1. Front-illuminated Images and Preprocessing

The closed-loop control system being updated by LAMOST utilizes a solution based on back-illuminated imaging. Within this system, the FPUs are illuminated from the spectrometer end by the back-illuminator to determine the optical fiber positions. Additionally, FPUs can be illuminated by the front-illuminator to capture their complete mechanical structures. The system employs six high-precision cameras, which collectively cover the entire focal plane. These cameras are positioned approximately 20 m from the focal plane. Each camera is equipped with an 800 mm focal length lens and a 7920 × 6004 pixel CMOS sensor. The image pixel size is 4.6 × 4.6 m, translating to 115 m on the focal plane or 1.18 arcseconds in the sky. For preliminary testing of the proposed methodology, only one camera was utilized in this study. The captured image is shown in Figure 1.

The captured image has a 24-bit bit depth, resulting in a pixel range of 0–$2^{24} - 1$, which is not optimal for subsequent processing. Standard algorithms typically operate on 8-bit images; therefore, it is necessary to convert the 24-bit image to an 8-bit format. A widely adopted approach to address this issue is grayscale stretching, which remaps the original pixel values to a new range. In this study, the pixel range [0, $2^{24} - 1$] is transformed to [0, 255]. The grayscale histogram of the original image is displayed in Figure 3. Notably, the captured image exhibits a significant difference between the maximum and minimum pixel values, exceeding 20,000. Nonetheless, the majority of pixel values predominantly fall within a narrower range of 4000–10,000. To ensure the inclusion of the white ceramic head information, pixels exceeding the value of 25,000 must also be considered. If these pixel values are directly scaled linearly to the [0, 255] range, the vast majority would be compressed into a range between 0 and 50. The resulting grayscale histogram is shown in Figure 4(b). The image generated after this mapping, as depicted in Figure 4(c), renders the mechanical structure of the Focal Plane Unit (FPU) almost entirely indiscernible, thereby obstructing further image analysis.

To mitigate this issue, logarithmic stretching is proposed as an effective solution. This technique employs a logarithmic mapping function to stretch pixel values in the low grayscale range, which are densely clustered, while compressing the pixel values in the high grayscale range, where values are more sparsely distributed. If the pixel value at position (x, y) is denoted as P , , the logarithmic transformation is applied using the following formula:

$$\bar{P}_{x,y} = A\log(P_{x,y} + 1)$$

where A is a constant scaling factor. The transformed pixel values are then rescaled to the range [0, 255]. To further refine the grayscale distribution and enhance the image contrast, only pixel values within a specified range are considered for this mapping. This selective approach prevents the lower pixel values, which tend to dominate due to the nature of logarithmic mapping, from occupying too many grayscale levels. As a result, this mitigates the compression of details in the low-intensity regions of the image. Figure 4(e) presents the grayscale histogram following this transformation, which still exhibits a significant imbalance in grayscale distribution. Consequently, as shown in Figure 4(f), the resulting image continues to obscure finer mechanical structures, demonstrating the limitations of this particular logarithmic mapping approach.

It has been observed that low-value background pixels, which typically contain minimal information, frequently occupy a significant portion of the grayscale range even after undergoing mapping. Meanwhile, the pixels that encode mechanical structure information are compressed into a very narrow range by the logarithmic function due to their higher values. This compression renders it challenging to distinguish between the mechanical structure and the background. To mitigate this issue, it is advisable to use the rapidly increasing part of the logarithmic function to map both the background and mechanical structure pixels. This can be achieved by introducing a constant parameter into the mapping formula, optimizing the contrast between these regions. The improved mapping formula is expressed as:

$$\bar{P}_{x,y} = A\log(P_{x,y} - b + 1)$$

Adding the constant b is analogous to performing a rightward shift of the logarithmic function as a whole, thereby enhancing the visibility of the black background and darker mechanical structures in regions where the function exhibits rapid changes. Given that the mapping is restricted to the pixel range (5000, 25,000), b must not exceed 5000; otherwise, negative pixel values will result after the mapping process. Experimental results have demonstrated that selecting b within the range (4500, 4900) yields the desired effect by improving the clarity of darker structures. Consequently, the median value b = 4700 is adopted in this work. Even if alternative values of b within this interval are adopted, the resulting image effects remain comparable. Moreover, as long as the image processing parameters are maintained consistently during both training and inference, the positioning accuracy will remain unaffected.

Under these conditions, the grayscale distribution of the image exhibits a well-balanced contrast, facilitating a clearer distinction between the background and the mechanical structure. The grayscale distribution has been effectively optimized, as demonstrated in Figures 4(h) and (i). It is important to emphasize that this method does not result in any loss of information. The value of b continues to represent the pixel value of the black background within the range of (0, 5000). During implementation, all pixel values less than b are assigned the value b before subtraction. This operation solely affects the background information and has no impact on the regions containing actual structural information.

2.2. Dataset

In this research, we utilize imagery captured under actual observational conditions to compose our training datasets. To train an AI model for fiber positioning, it is imperative to first establish a training dataset that includes paired data: front-illuminated fiber images alongside their precise positional coordinates. For each fiber image captured using front illumination, we employ the back-illumination technique to accurately ascertain the fiber's coordinates, which serve as the ground truth label for our dataset. To ensure exact correspondence between the fiber positions in the front- and back-illuminated images, both the fiber and the camera are held stationary during the acquisition process. Moreover, the images are captured in rapid succession to guarantee that the positional coordinates of the fibers remain consistent across both sets of images. The back-illumination method, known for its high accuracy, is widely utilized, thereby enhancing the reliability of the ground truth labels within our dataset. It is critical to emphasize that only the back-illuminated images are used for the dataset's formulation, enabling the AI model to effectively learn to deduce fiber positions from front-illuminated images. Once the model is trained, it can determine fiber positional coordinates solely from the front-illuminated images.

In the acquisition of fiber coordinates via back illumination, the two-dimensional Gaussian fitting method is typically employed. This method is predicated on the premise that the light emitted from the focal plane exhibits an ideal Gaussian intensity distribution (Dong & Wang 2012). Specifically, the intensity distribution across any perpendicular cross-section (x, y) of the light beam conforms to a Gaussian profile. The Gaussian intensity function is defined as:

$$I(x, y) = H \exp\left(-\frac{(x-x_0)^2}{2\sigma_1^2} - \frac{(y-y_0)^2}{2\sigma_2^2}\right)$$

Here, I(x, y) represents the intensity of the laser beam at the cross-section coordinates (x, y); H denotes the peak intensity of this cross-section; (x_0, y_0) specifies the central position of the spot; and σ_1, σ_2 are the standard deviations along the two orthogonal directions of the beam. Based on the functional form of this equation, the location of the peak intensity (H) is inferred to coincide with the central position of the spot. Thus, in this study, the position of the peak intensity is utilized to determine the central position of the spot.

Our dataset comprises more than 12,000 images, as shown in Figure 5(a). Of these, 60% constitute the training set, with the remainder divided between validation and test sets. Our final dataset consists of 50 × 50 pixel images that only contain the white ceramic header region of the FPU, which can be segmented utilizing the novel model developed by Professor Song's team at the LAMOST. Any cutting method can be used to reproduce our method, as long as there is a complete ceramic head in the cut image. We posit that this region sufficiently captures the majority of the necessary fiber position information, thereby obviating the need for a complete FPU image. To substantiate this hypothesis, we conducted a targeted comparative analysis, which demonstrated that employing the entire FPU as the dataset actually results in a significant decrease in accuracy. This outcome suggests that the full FPU image introduces extraneous information that detracts from model performance.

2.3. Network Architecture

Recent advances in AI have brought transformative progress across various fields. The development of deep learning techniques, particularly the extensive use of convolutional neural networks (CNNs), has significantly enhanced AI's ability to perform tasks such as image classification and object detection. These capabilities have proven invaluable in diverse domains, including medical imaging analysis and autonomous driving. Moreover, AI is increasingly being integrated into astronomical research, where it aids in optimizing telescope operations and data analysis. Image semantic segmentation technology was employed to estimate the initial orientation of the FPU, reducing the incidence of collisions (Zhou et al. 2021). Furthermore, object detection technology was utilized to develop an autofocus determination method tailored for the LAMOST. This method effectively enhances the closed-loop control system's performance under varying illumination conditions, both front and backlit (Zhou et al. 2022).

The evolution of neural network architectures for image processing has marked significant technological advancements, with milestones including AlexNet (Krizhevsky et al. 2012), ResNet (He et al. 2016), and the recent Swin Transformer (Liu et al. 2021). Each of these developments has enhanced the performance and broadened the applications of image processing technologies. In our method, we evaluated three distinct network architectures as potential backbones and compared their respective performance.

In 2012, AlexNet emerged as a pivotal advancement in deep learning for image processing. Utilizing a CNN architecture, it employed hierarchical convolution operations which substantially enhanced the accuracy of image classification. This breakthrough established AlexNet as a foundational model for deep learning applications. Introduced by Microsoft Research in 2015, the Residual Network (ResNet) provided an innovative solution to the issue of vanishing gradients in deep network training. ResNet's introduction of "skip connections" facilitated information retention and transmission across the network, addressing challenges associated with training deeper networks. This architecture not only increased the depth and performance of networks but also significantly mitigated the problem of overfitting. Due to its robust structure and superior performance, ResNet has found extensive applications in industrial settings, particularly in image recognition and object detection, and has become a benchmark in deep learning models. The Swin Transformer represents the latest innovation in image processing network structures. Integrating the self-attention mechanisms of the Transformer with the local processing benefits of convolutional networks, it hierarchically segments images into small patches and aggregates features across layers. This approach not only processes image features efficiently but also captures detailed local and global information. Demonstrating superior performance in various computer vision tasks, including image classification, object detection, and segmentation, the Swin Transformer offers higher accuracy and flexibility compared to traditional CNNs, positioning itself as the forthcoming standard in image processing networks.

Pre-trained models in deep learning are those initially trained on large-scale datasets, thereby acquiring preliminary weights that enable them to learn rich feature representations. These learned features are often generalizable and can effectively be transferred to various image processing tasks. The incorporation of this prior knowledge mitigates the demands on data and resources, enhances training efficiency, and accelerates the convergence rate of the final model. Furthermore, the parameters of pre-trained models, including learning rates, weight initialization methods, and the number of network layers, have been extensively optimized through large-scale experiments. This extensive pre-validation helps in avoiding common issues such as training instability, vanishing or exploding gradients, which are often encountered when training models from scratch. As a result, the training process and model performance remain robust and reliable. However, while these classic models often come with pre-trained versions, a limitation arises from their reliance on public datasets, which typically require a fixed input image size, most commonly 224 × 224 pixels. This constraint necessitates the interpolation of smaller images, such as 50 × 50 pixels, to fit the model's input size, which is suboptimal for high-precision tasks that require finer details. To address this issue, we trained ResNet18 from scratch to process 50 × 50 images, and based on it, designed smaller convolution kernels specifically for 50 × 50 images to make the growth of the receptive field more gradual, allowing the network to learn detailed local features layer by layer.

The architecture of the deep neural network under discussion employs a feature extractor, typically the network backbone, tasked with deriving high-dimensional semantic features from input imagery. Subsequently, these features undergo coordinate regression to yield the requisite positional outputs. We initially flattened the feature map obtained from the backbone network and employed a linear layer to transform the flattened features into coordinates. However, this approach did not meet the requisite accuracy standards. Consequently, we explored alternative methodologies to address this limitation. The critical phase of converting the high-dimensional features into precise coordinate predictions falls to the regression head, a component whose design significantly influences the overall accuracy of the model. This is particularly vital in tasks requiring sub-pixel precision, where effective coordinate prediction remains a challenging but essential aspect. Various methodologies for constructing regression heads are prevalent, including linear layer regression, heat map matching (Newell et al. 2016), and the Differentiable Spatial to Numerical Transform (DSNT) approach (Nibali et al. 2018). The linear layer regression technique, commonly implemented through a fully connected layer, enables direct prediction of target coordinates from features processed by the CNN. While straightforward and intuitive, this method may suffer from a lack of spatial generalization capability, and it often struggles with complex or highly variable input data structures. The heat map matching technique, extensively utilized in human pose estimation, offers an indirect means of keypoint location estimation by generating a heat map. On this map, high-value areas denote the probable locations of keypoints, aligning with their positions in the input image. However, the use of the argmax operation to estimate the predicted coordinate position inherently limits the output to integer values, which fails to meet the precision required for our application. Furthermore, this method demonstrates suboptimal performance when applied to low-resolution feature maps. Contrastingly, DSNT offers a sophisticated alternative that excels in handling low-resolution maps without accuracy degradation. This method, involving a fully differentiable layer, estimates keypoint coordinates by calculating a weighted average across the heat map. This approach not only facilitates numerical regression of coordinates but also ensures smooth gradient feedback across all pixels during inference, crucial for integrating DSNT within an end-to-end training framework. Crucially, DSNT's ability to compute non-integer coordinates and maintain efficacy across various resolutions marks a significant improvement over traditional heat map methods, rendering it particularly apt for tasks demanding high precision in coordinate prediction. Figure 6 is an example of a DSNT regression calculation.

The process underlying our method is illustrated in Figure 7, which outlines a two-stage framework. In the initial stage, we apply the segmentation technique devised by Professor Song's team at the LAMOST to divide the fiber focal plane image into several 50 × 50 pixel sub-images, each corresponding to a white ceramic fiber head. In the subsequent stage, we employ the backbone network combined with a regression head, as proposed in this study, to accurately predict the precise positions of the fiber heads.

3. Results

This section delineates the error metrics and result analysis derived from the aforementioned methodology, alongside a review of several comparative experimental outcomes. The error assessments are predicated on a dedicated test set comprising exclusively front-illuminated images, which were excluded from the training dataset. Utilizing this batch of 50 × 50 front-illuminated images, the trained model directly computed the fiber coordinates. The model's extrapolation capacity was then evaluated by measuring the discrepancy between the predicted coordinates and the actual ground-truth coordinates. This process underscores the model's effectiveness in generalizing from the training data to novel, unencountered images.

To improve fiber positioning accuracy, it is essential to first investigate which types of images yield the best performance. Figure 5 presents three different image datasets: a 50 × 50 pixel image of the ceramic head, a full view of the FPU, and an enlarged version of the ceramic head using cubic interpolation (suitable for pre-training purposes). To evaluate the effectiveness of each image type, we trained a ResNet18 model on each dataset and analyzed the accuracy of the predicted fiber coordinates. This approach aims to determine which image dataset provides the highest precision in fiber positioning.

To quantitatively assess the efficacy of our model on the test dataset, we employed two widely recognized metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The MAE quantifies the average magnitude of the absolute differences between the predicted values and the observed values, mathematically expressed as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where n is the total number of images in the test set, y represents the ith true coordinate, and ŷ denotes the ith predicted coordinate. This metric is particularly robust, as it assigns equal weight to all errors, thereby diminishing the influence of outliers. Conversely, RMSE is defined as the square root of the average of the squared differences between the predicted values and the observed values, calculated using the formula:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

RMSE is more sensitive to larger errors as it squares the deviations; this characteristic makes it especially useful for highlighting significant errors, albeit at the cost of increased sensitivity to outliers. Additionally, a statistical analysis of the error distribution revealed a close approximation to a Gaussian distribution, prompting the fitting of this distribution to our data. We subsequently derived the standard deviations for the x- and y-coordinates, providing crucial statistical insights into our positioning accuracy.

Table 1 shows the performance of the model trained on different image datasets. Our analysis reveals that the accuracy achieved using the full FPU image is significantly lower compared to the latter two methods and presents challenges in model convergence during training. This can be attributed to the predominance of critical fiber position data within the ceramic head, whereas the full

FPU image introduces excessive extraneous information that impairs learning efficiency and detracts from the model's predictive accuracy. This finding further confirms that utilizing a 50 × 50 pixel image does not result in any loss of positional information. As long as the image is consistently and accurately segmented, ensuring the complete preservation of the fiber ceramic ferrule within the frame, the stability of fiber detection can be reliably maintained.

Table 1 Performance Results of the ResNet18 Model Trained on Each Image Dataset

Image Type	MAE	MSE
50 × 50 image	**0.096**	**0.015**
Interpolation Resize	0.099	0.016
FPU image	0.120	0.023

Note. The fully connected (FC) layer is utilized as the regression head. The bold values indicate the highest precision achieved among all the schemes presented in the respective tables.

In assessing the performance of various established image processing models, we utilized a pre-trained backbone integrated with a linear layer to analyze their impact on the test dataset. The result is shown in Table 2. Transitioning from AlexNet to ResNet18 yielded a marked improvement in model efficacy. However, escalating the complexity of the model architecture, as seen with ResNet34 and the Swin Transformer, did not further enhance performance; rather, it exhibited a regression. We attribute this result primarily to the intrinsic characteristics of the image coordinate regression task and the available data volume. Unlike image classification, which heavily relies on high-level semantic features for object category identification, coordinate regression depends more on local image features such as edges, corners, and textures. These local features facilitate direct inference of coordinates through position-sensitive attributes like gradient direction. In this context, while the deeper ResNet34 model extracts higher-level semantic features more effectively than ResNet18, its increased depth also results in multiple convolutional and pooling operations. This, in turn, significantly reduces spatial resolution, blurring positional information and amplifying noise-related distortions. Similarly, the Swin Transformer model, which relies on a global attention mechanism for feature extraction, excels at capturing high-level semantic information but is less effective at learning local edge and texture features. As a result, its performance is inferior to the convolution-based ResNet18 in this specific regression task.

Furthermore, the dataset in this study is relatively small, aligning more closely with the VC dimension of ResNet18. In contrast, the larger parameter spaces of ResNet34 and the Swin Transformer make them more prone to overfitting when trained on limited data. Additionally, some random errors in the training set may exacerbate their susceptibility to overfitting. Consequently, these models are more likely to converge to local optima, diminishing their generalization capability and ultimately leading to suboptimal test performance.

For ResNet18, the best-performing pre-trained model in our experiments, we modified the initial convolutional kernel size and conducted pre-training from scratch using 50 × 50 images, which led to improved results. Notably, the ResNet18 backbone trained from scratch with the combination of the DSNT regression head yielded the highest accuracy across all performance metrics. Furthermore, Table 2 reports the prediction time for identifying all optical fibers (excluding object detection and image splitting) within the entire image. In tests conducted on a personal laptop, the ResNet18 architecture demonstrated superior performance in both accuracy and time efficiency. The integration of the DSNT regression layer not only enhanced accuracy but also did so with minimal impact on time consumption.

Table 2 Performance Outcomes of the Models

Model	MAE	MSE	Param	Time
AlexNet(Pretrain)	0.115	0.021	2.1 s	-
ResNet18(Pretrain)	0.096	0.015	5.5 s	-
ResNet34(Pretrain)	0.102	0.018	11.1 s	-
Swin Transformer v2(Pretrain)	0.108	0.019	14.6 s	-
ResNet18(Scratch)	0.093	0.014	5.0 s	-
ResNet18+DSNT (Scratch)	**0.087**	**0.012**	5.3 s	-

Note. Pretrain refers to models initialized with pretrained weights, whereas Scratch denotes models trained from the beginning without the use of any pre-trained parameters. Time is the duration to predict fiber coordinates on a 7920 × 6004 focal plane image, excluding target detection and image splitting. Param denotes the model's parameter count. The bold values indicate the highest precision achieved among all the schemes presented in the respective tables.

The following section provides a comprehensive analysis of the error distribution associated with the prediction outcomes obtained from the ResNet18+DSNT model, as illustrated in Figure 8. Let x d represent the x-coordinate predicted by the model and x c denote the x-coordinate derived via the back-illumination method. The error in the x-coordinate is then defined as Δx = x d − x c , with Δy representing the analogous error in the y-coordinate. Since both Δx and Δy are assumed to conform to a Gaussian (normal) distribution, the overall error, which is given by D = √(Δx^2 + Δy^2), conforms to a Rayleigh distribution. The probability density function (PDF) of the Rayleigh distribution is given by:

$$f(d) = \frac{d}{\sigma^2} \exp\left(-\frac{d^2}{2\sigma^2}\right)$$

where σ is the scale parameter related to the standard deviation of the Gaussian-distributed components Δx and Δy. The expected value and standard deviation of Δd can be expressed in terms of σ, with the expectation given by:

$$E[d] = \sigma\sqrt{\frac{\pi}{2}}$$

and the standard deviation by:

$$\sigma_d = \sigma\sqrt{\frac{4-\pi}{2}}$$

The figure depicts the distributions of Δx, Δy, and Δd, alongside their respective Gaussian and Rayleigh fits. Figure 8 presents the statistical distribution of measurement errors across all fibers in the test dataset, encapsulating the aggregate error distribution resulting from multiple shots. This figure effectively illustrates the overall performance and reliability of the fibers throughout the dataset. The calculated standard deviations for these distributions are 0.12 and 0.11 pixels, respectively, which sufficiently satisfy the accuracy requirements of the project.

Figure 9 illustrates the distribution of error vectors for all optical fibers situated on a single focal plane within the test set. Each sub-image is derived from photographs of the focal plane captured subsequent to the moving of the optical fibers. Figure 10 presents a statistical diagram that corresponds to the sub-images in Figure 9, depicting the distribution of errors associated with each. For example, Figures 9(a) and 10(a) are derived from the same shooting of optical fibers, respectively, which are the error vector distribution map and error statistics map of the method for this shooting. In Figures 9 and 10, the first sub-image depicts all optical fibers at their home positions, whereas the subsequent sub-images illustrate the fibers at various offset positions. Notably, our methodology demonstrates marginally enhanced positioning accuracy for fibers in the home position compared to those in offset positions. Although the positioning of optical fibers may sometimes exhibit slightly larger errors in certain instances, likely due to random external influences, the overall precision remains consistently high across multiple measurements (as depicted in Figure 8). This level of accuracy satisfies the closed-loop system requirements for LAMOST fiber positioning and underscores the robustness of our approach.

4. Discussion

In Zhou's study (Zhou et al. 2022), it was suggested that the camera's orientation could lead to systematic deviations between the front and back images. However, the error map we generated does not exhibit such a pattern, suggesting that this deviation can potentially be captured and accounted for by the model. Our observed errors appear to be more random and irregular, showing no discernible correlation with the position of the optical fibers on the focal plane. We also explored the use of the CoordConv method (Liu et al. 2018), originally proposed to address the issue of convolutional networks being insensitive to spatial position, by incorporating approximate positional information of the optical fiber on the focal plane as input for model training. Specifically, after gridding the entire image, the approximate location, as identified by the first stage of target detection, was utilized to determine the grid in which the optical fiber resides. This information was then integrated into the model through the CoordConv layer during training. However, this approach did not result in improved accuracy, indicating that the source of error is likely unrelated to the camera's angle or the position of the optical fiber on the focal plane.

In Figure 9, the distribution of single-fiber measurement results appears largely random. However, the statistical analysis depicted in Figure 10 reveals a small numerical offset in the error, as indicated by the values of the variable   representing the Δx and Δy offsets. Notably, these offsets appear to be randomly distributed around the central point, with no significant bias in the aggregated results of multiple measurements (as shown in Figure 8). This pattern suggests that the observed deviations may stem from stochastic variables in the single image acquisition rather than the AI methodology employed. We suggest that during the dataset construction phase, even though the fiber and the camera maintain a fixed position for consecutive front and back-illuminated shooting, subtle shifts in the fiber's position in the image may occur due to extrinsic factors such as camera shake or atmospheric turbulence, leading to minor, random displacements.

To test this hypothesis, the experimental protocol was designed to replicate the dataset construction conditions: the fiber and the camera were held stationary for a brief period while capturing two successive back-illuminated photographs and using the same back-illuminated fiber detection algorithm to determine the fiber position coordinates. The decision to conduct the error analysis using two consecutive back images, rather than front and back images, stems from the complexity of isolating errors arising from hardware factors and algorithm factors between front and back exposures. These non-algorithmic influences make it challenging to determine whether observed discrepancies originate from the algorithm itself or from external factors. To address this issue, we replicate the experimental conditions of sequential front and back imaging by capturing two consecutive back images. Using the same positioning algorithm based on back illumination, we extract the respective coordinates, thereby eliminating algorithmic variations as a potential source of error. Consequently, any differences in the computed coordinates between the two consecutive back images can be attributed solely to non-algorithmic factors, such as hardware limitations, environmental conditions, and temporal fluctuations. The error vector diagram, presented in Figure 11, corroborates our hypothesis by demonstrating a similar error distribution, thereby affirming that these random errors likely arise from variations in the imaging process rather than the AI algorithm itself.

We believe that the accuracy of alignment could potentially be enhanced through the development of algorithms designed to establish multiple fixed reference points on the focal plane in the front-illuminated images. These reference points would facilitate precise coordinate alignment post-imaging. Nevertheless, it is imperative that the precision of these fixed points substantially surpasses the offsets ( ) illustrated in Figure 10, as insufficient accuracy could prove detrimental rather than beneficial. However, it is important to note that even in the absence of these enhancements, the AI method deployed herein satisfies the stringent accuracy requirements mandated by the LAMOST fiber detection system. This performance is commensurate with that achieved via back-illuminated techniques.

5. Conclusions

To achieve greater fiber positioning accuracy, LAMOST is currently developing a closed-loop control fiber detection system, a pivotal component of the overall closed-loop control framework. Unlike the back-illuminated imaging technique, which solely illuminates fibers from the spectrometer end, the front-illuminated imaging captures comprehensive information from the FPU, offering an alternative method to calculate fiber positions. This study introduces an advanced fiber detection approach that leverages cutting-edge AI technology based on front-illuminated images captured by the fiber monitoring camera. Specifically, object detection techniques are employed to extract the white ceramic ferrule image encompassing the fiber. The AI model is subsequently trained using data collected from actual observations, with the fiber positions determined by back-illuminated imaging serving as ground-truth labels. This enables the AI model to predict fiber coordinates from the front-illuminated images. A subset of the observational data, which was excluded from the training process, was utilized to evaluate the extrapolation capability of the AI model. The experimental results demonstrated that the ResNet18 architecture yielded the best performance in terms of both accuracy and processing time. Notably, the DSNT regression head further enhanced the model's accuracy, achieving a precision level of 0.11 pixels—sufficient to meet LAMOST's fiber positioning accuracy requirements. In contrast to the back-illuminated system, which introduces additional observation overheads and light pollution within the spectrograph during nighttime operations, the proposed front-illuminated method effectively mitigates these issues. As a result, this technique not only fulfills the stringent accuracy standards set by LAMOST but also presents a promising solution for implementation in the future closed-loop control fiber positioning system.

Acknowledgments

We extend our sincere appreciation to the National Natural Science Foundation of China (NSFC) for providing the financial support that made this research possible. Specifically, this work was supported by NSFC under grant Nos. 12073046 and 12261141689. The funds provided were instrumental in facilitating the experiments and data analysis. Guo Shou Jing Telescope (the Large sky Area Multi-Object fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

References

Baltay, C., Rabinowitz, D., Besuner, R., et al. 2019, PASP, 131, 065001
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
Dong, H., & Wang, L. 2012, Optik, 123, 2148
Drass, H., Vanzi, L., Torres-Torriti, M., et al. 2016, Proc. SPIE, 9908, 99088E
Fisher, C., Morantz, C., Braun, D., et al. 2014, Proc. SPIE, 9151, 91511Y
He, K., Zhang, X., Ren, S., & Sun, J. 2016, Deep Residual Learning for Image Recognition, in 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV: IEEE), 770
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, Commun. ACM, 60, 84
Liu, R., Lehman, J., Molino, P., et al. 2018, arXiv:1807.03247
Liu, Z., Lin, Y., Cao, Y., et al. 2021, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, in 2021 IEEE/CVF Int. Conf. Computer Vision (ICCV) (Montreal, QC: IEEE), 9992
Montgomery, D., Atkinson, D., Beard, S., et al. 2016, Proc. SPIE, 9908, 99081R
Newell, A., Yang, K., & Deng, J. 2016, Stacked Hourglass Networks for Human Pose Estimation, in Computer Vision—ECCV 2016, Vol. 9912, ed. B. Leibe, et al. (Cham: Springer International Publishing), 483
Nibali, A., He, Z., Morgan, S., & Prendergast, L. 2018, Numerical Coordinate Regression with Convolutional Neural Networks, arXiv:1801.07372
Schubnell, M., Ameel, J., Besuner, R. W., et al. 2016, Proc. SPIE, 9908, 99081Q
Wang, S.-Y., Chou, R. C. Y., Huang, P.-J., et al. 2016, Proc. SPIE, 9908, 99083Y
Xing, X., Zhai, C., Du, H., et al. 1998, Proc. SPIE, 3352, 839
Zhou, M., Lv, G., Li, J., et al. 2021, PASP, 133, 115001
Zhou, M., Zhang, Y., Lv, G., et al. 2022, RAA, 22, 065004
Zhou, Z., Liang, J., Duan, S., et al. 2022, PASP, 134, 025001
Zhou, Z., Wang, J., Hu, H., et al. 2018, Proc. SPIE, 10700, 77

Note: Figure translations are in progress. See original paper for figures.