

SAM-Canny: A Zero-Shot Segmentation Guided Adaptive Edge Detection Algorithm for High-Precision Visual Measurement of CuCrZr frames

yuheng Wang

Beijing Information Science & Technology University

yabing Wang

Institute of High Energy Physics

yuhai Gu

guyuhai@bistu.edu.cn

Beijing Information Science & Technology University

weiwei Zhang

Institute of High Energy Physics

Research Article

Keywords: Machine vision, Edge detection, Error compensation, SAM3, CuCrZr frames

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8344455/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

SAM-Canny: A Zero-Shot Segmentation Guided Adaptive Edge Detection Algorithm for High-Precision Visual Measurement of CuCrZr frames

Yuheng Wang^{1†}, Yabing Wang^{2†}, Yuhai Gu^{1*}, Weiwei Zhang^{2*}

¹School of Mechanical and Electrical Engineering, Beijing Information Science and Technology University, 12 Xiaoying East Road, Beijing, 100192, China.

²Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Beijing, 100049, China.

*Corresponding author(s). E-mail(s): guyuhai@bistu.edu.cn;
zhangww@ihep.ac.cn;

Contributing authors: 13730237366@163.com; ybwang@ihep.ac.cn;

[†]These authors contributed equally to this work.

Abstract

The CuCrZr frame is a key focusing component of the Beijing Huairou High Energy Photon Source. To improve the measurement accuracy and automation of its diameter, this paper proposes a high-precision visual measurement framework that integrates zero-shot segmentation with adaptive edge detection. The framework first employs the Segment Anything Model 3 (SAM3) to robustly extract the target region without any training. Next, the SAM-Canny adaptive edge detection operator is introduced, which searches for the optimal dilation kernel size for the Otsu-Canny algorithm via progressive morphological dilation and automatically determines double thresholds using the Otsu method to extract edge pixels. Subpixel edge localization is then achieved using RANSAC combined with Zernike moments. Finally, a whale optimization algorithm-enhanced radial basis function (WOA-RBF) neural network is constructed to build a multidimensional error compensation model based on the grayscale features of different material regions segmented by SAM, effectively correcting measurement deviations caused by illumination variations. The effectiveness of the proposed method is demonstrated through diameter measurement and error compensation

experiments. In a generalization test on five new CuCrZr frames, the method maintains single-micrometer-level errors (2.4–6.3 μm), significantly outperforming traditional CuCrZr frame inspection methods, which exhibit errors up to 32.1 μm .

Keywords: Machine vision, Edge detection, Error compensation, SAM3, CuCrZr frames

1 Introduction

As a world-class large scientific facility under construction, the Beijing Advanced Photon Source (HEPS) in Huairou is expected to generate extremely high-brightness X-rays, which render conventional optical components unable to withstand the high thermal load and intense radiation environment. Diamond compound refractive lenses (CRLs), owing to their excellent thermal stability and radiation resistance, have become the key components for achieving high-flux X-ray focusing on HEPS beam-lines [1]. In practice, CRLs are composed of numerous diamond lenses precisely arranged in V-shaped grooves to achieve continuous focusing, which imposes stringent requirements on the central alignment of the lens structures. However, the ultrahard nature of diamond makes high-precision machining extremely challenging, and ensuring consistent lens diameters is difficult [2]. To address this, a “mount-first, then process” approach is commonly adopted internationally, whereby diamond lenses are first pressed into the centers of easily machinable cylindrical CuCrZr frames (as shown in Fig. 1), followed by laser machining of the lens structures within the diamond. This procedure not only improves the consistency of diamond lens diameters but also significantly enhances the alignment accuracy of the lens machining centers [3–7].

Currently, the diameter measurement and quality inspection of CuCrZr frames mainly rely on manual operations using contact or non-contact instruments. While suitable for small-batch inspection, manual measurements are inefficient and prone to errors when faced with the large-scale production demands of HEPS. With the advancement of machine vision, non-contact visual measurement has become an important approach for industrial dimensional inspection [8]. Existing studies have reported sub-pixel edge detection and high-precision aperture measurement using various improved Canny-based or model-driven methods. However, most of these approaches rely on manually tuned parameters and are often developed for specific types of components. Their generalization ability is limited, particularly under varying illumination and material conditions. For example, Jun Guo et al. proposed a sub-pixel Canny–Zernike method with improved adaptive thresholds for high-precision detection of flange apertures [9]; Libin et al. combined wavelet denoising with modified Canny and Sobel operators to improve edge detection of shaft components [10]; Yuntao Fang et al. developed a model-driven sub-pixel measurement method by optimizing autofocus and threshold selection strategies for ring-shaped workpieces [11]; Xuelin Zhang et al. designed a direction-adaptive Canny operator with chain-length filtering to effectively reduce broken and spurious edges [12]. However, these methods are typically

application-specific, with limited generalization capability, and still require manual tuning of algorithm parameters, making them unsuitable for direct application to CuCrZr frames measurement.

The diameter measurement of CuCrZr frames involves three main challenges. The first challenge is the reliable extraction of the region of interest (ROI). Traditional ROI extraction methods rely on manually designed rules, which often require substantial engineering effort and generalize poorly across varying backgrounds or illumination conditions. These methods may fail when strong reflections or complex textures are present. Although deep learning-based graph-cut techniques can improve ROI extraction, they require extensive training data and are difficult to deploy in flexible industrial environments.

The second challenge concerns the selection of thresholds in the Canny operator. Although Canny generally provides better performance than other traditional edge detectors, its high and low thresholds must be tuned for different imaging conditions. Existing adaptive variants exhibit various limitations: Otsu-based methods assume strong foreground-background separation and may fail on low-contrast images; percentile-based methods still require manually selecting percentiles; gradient-based local adaptive approaches incur high computational cost; and deep learning-based edge detectors require substantial training resources [13].

The third challenge lies in illumination-induced measurement errors. Variations in lighting intensity affect grayscale values through both noise and sensor saturation. Insufficient illumination introduces Poisson noise [14], whereas excessive illumination may cause gain saturation and detail loss [15]. These effects introduce systematic diameter deviations. Existing compensation strategies typically use a one-dimensional model relating illumination to measurement error [16]. However, such models do not account for the different reflectance properties of multiple materials in the scene, limiting their ability to accurately correct illumination-dependent errors.

To overcome the limitations of existing methods, we develop a unified measurement framework that integrates zero-shot segmentation, adaptive edge extraction, and data-driven error compensation. SAM3 is employed to isolate the region of interest in a training-free manner [17]. The SAM-Canny operator is then introduced to achieve illumination-robust edge detection by combining SAM-based structural cues with adaptive threshold selection. The detected edges are refined through RANSAC filtering and Zernike-moment subpixel localization [18, 19]. Furthermore, a multi-dimensional compensation model utilizing grayscale features from different material regions is constructed to account for illumination-induced systematic errors.

The remainder of this paper is organized as follows: Section 2 introduces the zero-shot ROI extraction based on SAM and the SAM-Canny operator, and details the construction of the multidimensional error compensation model; Section 3 presents the experimental setup, results, and performance analysis; Section 4 concludes the paper.

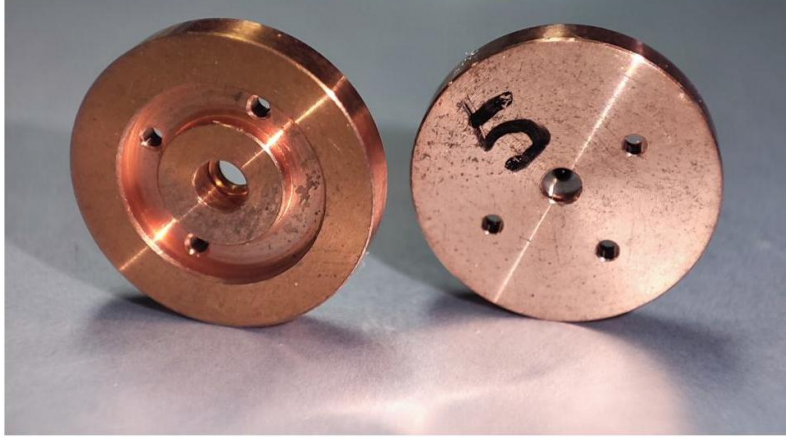


Fig. 1 Copper-Chromium-Zirconium frames

2 Theoretical framework

2.1 Segment Anything Model 3

The Segment Anything Model 3 (SAM3) is a prompt-driven segmentation framework capable of producing object masks without task-specific training. Given an input image $I \in \mathbb{R}^{H \times W \times 3}$ and a prompt P , the model generates a binary mask $M \in \{0, 1\}^{H \times W}$:

$$M = \text{SAM3}(I, P), \quad (1)$$

where H and W denote the spatial resolution of the image. SAM3 adopts a general encoder-decoder architecture in which a multi-scale encoder extracts visual features and a prompt-guided decoder synthesizes the final mask:

$$M = D(E(I), P). \quad (2)$$

For a point prompt p , the prompt is represented as a one-hot map P_p , and the mask generation becomes

$$M = D(E(I), P_p). \quad (3)$$

The primary advantage of SAM3 lies in its zero-shot segmentation capability, enabling reliable extraction of regions of interest (ROIs) across diverse object shapes and imaging conditions. This property makes SAM3 well suited for coarse edge localization and ROI delineation in subsequent measurement stages.

To evaluate the robustness of SAM3 under varying illumination, segmentation results obtained under extreme lighting conditions are shown in Fig. 2. SAM3 consistently produces accurate masks across both high- and low-illumination scenarios, demonstrating strong resilience to intensity fluctuations in the CuCrZr frame imaging setup.

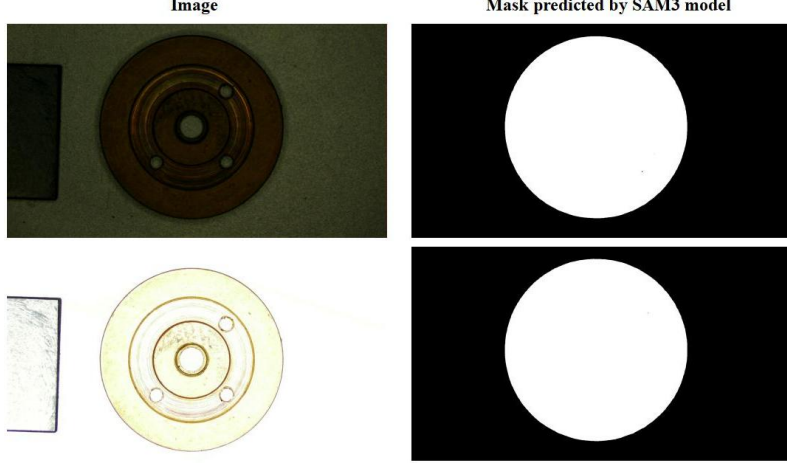


Fig. 2 Segmentation results of the SAM3 model under varying illumination.

2.2 SAM-Canny

After obtaining the coarse edge mask from SAM segmentation, this study proposes an adaptive and parameter-free edge refinement method. Let $I(x, y)$ denote the grayscale input image and M_{edge} the coarse SAM edge set. The Sobel gradients of the input image are computed to obtain the gradient magnitude:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}, \quad G_x = I * S_x, \quad G_y = I * S_y. \quad (4)$$

The gradient values of the coarse SAM edges are collected as

$$\mathcal{A} = \{ G(x, y) \mid (x, y) \in M_{\text{edge}} \}. \quad (5)$$

To construct an edge-adjacent region suitable for threshold estimation, we adaptively determine the dilation kernel size used to expand the SAM mask. Discrete kernel sizes (e.g., 3×3 , 5×5 , 7×7) are evaluated. Each dilation step produces a thin ring-shaped region surrounding the SAM boundary:

$$\text{ROI}_k = D_k(M_{\text{edge}}) \setminus M_{\text{edge}}, \quad (6)$$

where $D_k(\cdot)$ denotes dilation using a $k \times k$ structuring element.

For each candidate region ROI_k , the gradient values are collected:

$$\mathcal{R}_k = \{ G(x, y) \mid (x, y) \in \text{ROI}_k \}. \quad (7)$$

From this set, the histogram is computed and the between-class variance is obtained according to Otsu's formulation:

$$\sigma_B^2(k) = w_0(k)(\mu_0(k) - \mu_T(k))^2 + w_1(k)(\mu_1(k) - \mu_T(k))^2. \quad (8)$$

The optimal dilation kernel size k^* is selected using a progressive dilation search:

$$k^* = \arg \max_k \sigma_B^2(k). \quad (9)$$

The final ring-shaped ROI is then defined as

$$\text{ROI} = D_{k^*}(M_{\text{edge}}) \setminus M_{\text{edge}}. \quad (10)$$

Within this ROI, the high threshold for Canny is obtained directly using Otsu's method:

$$T_h = \text{Otsu}(\mathcal{R}_{k^*}). \quad (11)$$

The low threshold is estimated from the coarse-edge gradient set:

$$T_\ell = \max(\min(\mathcal{A}, P_p(\mathcal{A}), \varepsilon), \quad (12)$$

where p is a small percentage (typically between 2 and 5) and ε is a small positive constant.

Finally, Canny hysteresis thresholding is applied within the ROI using (T_h, T_ℓ) . Pixels with $G(x, y) \geq T_h$ are marked as strong edges, and weak edges with $T_\ell \leq G(x, y) < T_h$ are retained if they are 8-connected to any strong edge:

$$\|(x, y) - (u, v)\|_\infty = 1 \text{ and } G(u, v) \geq T_h \implies (x, y) \in \mathcal{E}, \quad (13)$$

where $\mathcal{E}$ denotes the final refined edge set.

After hysteresis thresholding, the Canny operator yields a preliminary refined edge set

$$\mathcal{E} = \{(x, y) \mid G(x, y) \geq T_\ell \text{ and retained through hysteresis}\}. \quad (14)$$

Although the SAM-guided thresholds suppress most background responses, a small number of geometric outliers may still occur. To ensure structural consistency, a circle-model RANSAC procedure is applied to $\mathcal{E}$. In each iteration, three points are randomly sampled to form a candidate circle hypothesis (c_x, c_y, r) , and its consensus set is computed as

$$\mathcal{I} = \{(x, y) \in \mathcal{E} \mid |\sqrt{(x - c_x)^2 + (y - c_y)^2} - r| < \tau\}, \quad (15)$$

where τ denotes the radial tolerance. The hypothesis with the largest consensus set is selected as the optimal inlier set $\mathcal{I}^*$.

To further enhance localization accuracy, Zernike moment-based subpixel refinement is performed on each inlier. For an inlier pixel (x, y) , the Zernike coefficients within a local neighborhood $\Omega(x, y)$ are evaluated as

$$Z_{nm}(x, y) = \iint_{\Omega(x, y)} I(u, v) V_{nm}^*(u - x, v - y) du dv, \quad (16)$$

where V_{nm} denotes the Zernike basis. A continuous edge response is subsequently reconstructed, and the subpixel location is estimated by

$$(x_s, y_s) = \arg \max_{(u,v)} \left| \sum_{n,m} Z_{nm}(x, y) V_{nm}(u - x, v - y) \right|. \quad (17)$$

The complete set of subpixel-refined inliers is denoted by

$$\mathcal{I}_s = \{(x_s, y_s) \mid (x, y) \in \mathcal{I}^*\}. \quad (18)$$

Finally, a geometric least-squares circle fitting is applied to $\mathcal{I}_s$. The optimal circle parameters are obtained by minimizing the radial residuals:

$$(c_x^*, c_y^*, r^*) = \arg \min_{c_x, c_y, r} \sum_{(x_s, y_s) \in \mathcal{I}_s} \left(\sqrt{(x_s - c_x)^2 + (y_s - c_y)^2} - r \right)^2. \quad (19)$$

The corresponding diameter is computed as

$$d^* = 2r^*. \quad (20)$$

Through the integration of RANSAC-based outlier rejection, Zernike moment subpixel refinement, and least-squares geometric fitting, the proposed SAM-Canny pipeline achieves robust and highly precise diameter estimation.

All thresholds and the dilation kernel size are adaptively determined from mask geometry and local gradient statistics, enabling fully automatic, zero-parameter precise edge localization.

2.3 Multi-Dimensional Grayscale Feature Input and WOA-RBF Modeling

In machine vision online measurement systems, the error Δ in workpiece edge extraction is mainly caused by fluctuations in ambient illumination E . The physical mechanism of the measurement error caused by variable illumination is that changes in light intensity lead to changes in the image grayscale distribution, which in turn affects the accuracy of edge localization. According to photometry and imaging theory, the grayscale value $G(x, y)$ captured by the camera at the workpiece surface point (x, y) is determined by the ambient illumination E , the workpiece surface's Bidirectional Reflectance Distribution Function (BRDF) $f_r(L, V)$, and the imaging system parameters C (such as exposure time T , gain K , etc.). The simplified imaging equation is expressed as:

$$G(x, y) = K \cdot T \cdot \int_{\Omega} E(\mathbf{L}) \cdot f_r(\mathbf{L}, \mathbf{V}) \cdot \cos(\theta_L) d\Omega + G_{\text{noise}} \quad (21)$$

Existing research (e.g., the method by Su et al.) tends to use the ambient illuminance meter reading E_s as the sole input feature, assuming $\Delta = g(E_s)$. However, this simplified modeling approach has serious drawbacks: different materials possess vastly

different BRDFs f_r , meaning that even if E_s is constant, the image grayscale distribution $G(x, y)$ on the surface of different materials will vary significantly. The model input E_s fails to distinguish between grayscale changes caused by material variation (Δf_r) and those caused by illumination variation (ΔE), leading to extremely poor generalization ability.

To address the limitations of the single illumination input E , this study proposes using SAM segmentation results to extract a semantically meaningful multi-dimensional grayscale feature vector $\mathbf{X}$ as the input for the error prediction model. First, the SAM model is used to perform high-precision segmentation of the image containing the workpiece, obtaining precise masks M_k for the workpiece body, background, and potentially adjacent reference areas. Subsequently, the grayscale values of all pixels within these segmentation masks M_k are statistically analyzed to extract the average grayscale value $\bar{G}_k$, for each region, which serves as a representative of the optical response of that specific material (or area) under the current lighting conditions. Finally, the average grayscale values of N critical regions are combined into the multi-dimensional input feature vector $\mathbf{X}$:

$$\mathbf{X} = [\bar{G}_1, \bar{G}_2, \dots, \bar{G}_N]^T \quad (22)$$

This vector $\mathbf{X}$ vectorially encodes the light intensity information of different material surfaces under the current illumination, directly reflecting the coupled influence of $E \cdot f_r$, which establishes a stronger correlation between $\mathbf{X}$ and the edge localization error Δ , resolving the issue of single E input's inability to adapt to multi-material workpieces.

Subsequently, we employ the Whale Optimization Algorithm-Optimized Radial Basis Function Neural Network (WOA-RBF) to establish the nonlinear mapping relationship between the multi-dimensional grayscale feature vector $\mathbf{X}$ and the visual measurement error Δ :

$$\Delta = f_{\text{WOA-RBF}}(\mathbf{X}) \quad (23)$$

The theoretical basis for selecting WOA-RBF is that the RBF neural network possesses properties of best approximation capability, simple structure, and fast convergence, making it suitable for dealing with nonlinear, small-sample regression problems in visual measurement. Given that RBF network performance is highly sensitive to parameter selection (center points c_i , spread constants σ_i , weights w_i), the Whale Optimization Algorithm (WOA) is introduced. WOA is utilized for its strengths in strong search capability and fast convergence, enabling the efficient finding of the optimal parameter combination for the RBF network globally. This is expected to significantly improve the model's accuracy, stability, and generalization ability to achieve precise compensation for micrometer-level measurement errors.

In this study, the neural network has five inputs, as illustrated in Fig. 3. These inputs are the average gray value of the standard gauge block, the average gray value of the outer ring of the CuCrZr frames, the average gray value of the middle ring of the CuCrZr frames, the average gray value of the inner ring of the CuCrZr frames, and the average gray value of the backplane.

The overall flowchart of the proposed measurement algorithm is illustrated in Fig. 4. The process begins with image acquisition, followed by SAM3-based ROI extraction, adaptive edge detection using SAM-Canny, and concludes with error compensation via the WOA-RBF model.

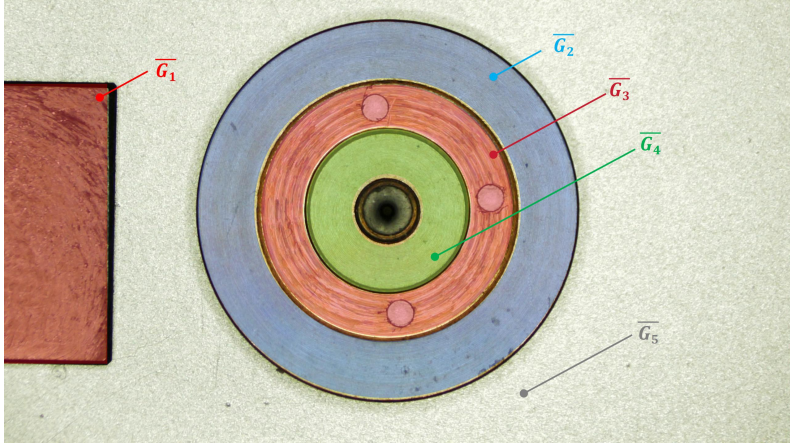


Fig. 3 Input of neural network

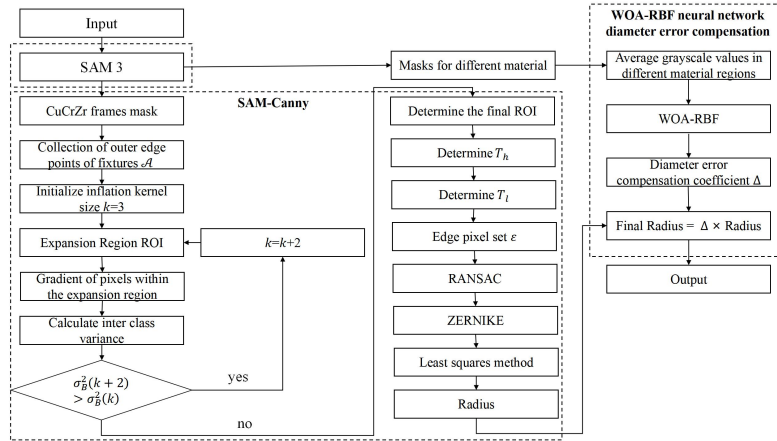


Fig. 4 Algorithm flowchart

3 Diameter Measurement Experiment

3.1 Measurement systems and calibration

The experimental platform was constructed based on a high-resolution CCD imaging system. The imaging module employs a GP-660V microscopic CCD camera with a

resolution of 1920×1080 pixels, rigidly mounted on a precision focusing mechanism to ensure image sharpness and visibility of small-diameter features. Illumination is provided by a D-100 ring light source. The experiment-control computer is equipped with an Intel Core i5-12400F processor and 16 GB of memory, running Windows 10 (64-bit), and is responsible for image acquisition, calibration, and subsequent algorithmic processing.

To obtain high-precision geometric parameters of the camera, a 7×7 circular calibration target was used, where each circular marker has a diameter of 0.625 mm and an inter-center spacing of 1.25 mm. Subpixel-accurate pixel coordinates of the calibration circles were extracted and matched with their corresponding three-dimensional physical coordinates to solve the intrinsic and extrinsic camera parameters and compensate for lens distortion, thereby ensuring the geometric accuracy of diameter measurement.

The ground-truth diameters were measured using a KEYENCE VK-X3000 laser confocal microscope. The system is equipped with a $20\times$ objective lens and provides a spatial resolution of 686.688 nm/pixel, enabling nanometer-level measurement accuracy. This precision significantly exceeds that of the CCD imaging system, ensuring the reliability of the reference values used for performance evaluation.

3.2 Diameter measurement dataset

A chromium–zirconium–copper (CuCrZr) frames was employed as the experimental specimen to construct the dataset required for diameter measurement. The frame was positioned on the baseplate of the CCD imaging system, after which the camera focus was precisely adjusted to be consistent with that of the circular calibration target, ensuring uniform imaging scale across all acquisitions. The ring light source was subsequently driven through a controlled illumination sweep, during which its intensity was incrementally increased from low to high levels. Throughout this illumination progression, the CCD camera continuously acquired images under fixed optical and exposure settings, thereby generating a temporally ordered sequence that encompasses a broad range of lighting conditions.

Following data acquisition, all images underwent a manual quality-screening process to remove defocused frames, motion-blurred samples, and redundant observations. A total of 284 high-quality CuCrZr frames images were retained. The resulting dataset captures systematic illumination variability while preserving geometric fidelity and image sharpness, thereby closely approximating the operational characteristics encountered in practical industrial inspection environments. As such, it provides a reliable and representative foundation for rigorously evaluating the accuracy and robustness of the proposed diameter measurement algorithm.

3.3 Statistical indicators

To quantitatively evaluate the accuracy and stability of the proposed diameter measurement algorithm, this study adopts several metrology-oriented statistical indicators, including the Maximum Error, Bias, Standard Deviation (STD), and the Within-Tolerance Rate (WTR). These indicators jointly characterize the extremum deviation,

systematic tendency, fluctuation level, and tolerance compliance of the measurement results. Let D_i denote the ground-truth diameter, $\hat{D}_i$ the estimated diameter, $e_i = \hat{D}_i - D_i$ the measurement error, and N the total number of samples. The definitions of the evaluation metrics are described as follows.

The Maximum Error quantifies the worst-case deviation between the estimated and true diameters:

$$E_{\max} = \max_{i=1, \dots, N} |e_i|. \quad (24)$$

The Bias characterizes the systematic offset of the algorithm, reflecting whether the diameter tends to be consistently overestimated or underestimated:

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N e_i. \quad (25)$$

The Standard Deviation (STD) measures the dispersion of the measurement errors around the bias, thereby indicating the stability and repeatability of the algorithm:

$$\text{STD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i - \text{Bias})^2}. \quad (26)$$

To evaluate whether the measurements satisfy the engineering accuracy requirement, the Within-Tolerance Rate (WTR) is defined as the proportion of samples whose absolute error is within a predefined tolerance threshold τ (set to $15 \mu\text{m}$ in this work):

$$\text{WTR} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(|e_i| \leq \tau) \times 100\%, \quad (27)$$

where $\mathbb{I}(\cdot)$ is the indicator function, which equals 1 if the condition is satisfied and 0 otherwise.

These statistical indicators provide a comprehensive perspective on measurement performance, enabling a rigorous assessment of the algorithm's accuracy, systematic deviation characteristics, stability, and compliance with precision engineering tolerances.

3.4 Experimental Results and Analysis of Diameter Measurement

In traditional industrial component vision measurement tasks, the method combining Canny edge detection with Zernike moment-based subpixel refinement is widely used. In related studies [9–11, 20], this method is typically employed as a baseline algorithm to validate the measurement accuracy of newly proposed methods. For comparison, this work implements a conventional Canny–Zernike baseline measurement pipeline: first, the grayscale image is filtered using a median filter to reduce noise; then, the initial boundary of the Cr–Zr–Cu frame is extracted using fixed-threshold

Canny edge detection, where the threshold is selected as the optimal value through multiple experiments; subsequently, Zernike moments are used for subpixel edge localization refinement; finally, the frame diameter is calculated via least-squares geometric fitting. This algorithm serves as the reference baseline for evaluating the accuracy of the proposed method.

The experimental results presented in Table 1 demonstrate the significant advantages of the proposed SAM-Canny method over the traditional Canny operator for diameter measurement of CuCrZr frames. The SAM-Canny method achieved a maximum error (Emax) of 18.15 μm , representing a 35.8% reduction compared to the 28.31 μm of the conventional approach. More importantly, the systematic bias was dramatically reduced from +20.61 μm to -6.49 μm , indicating effective mitigation of measurement inaccuracies.

The Within-Tolerance Rate (WTR) further highlights the improvement. The proposed SAM-Canny method achieves a WTR of 92.9% under a 15 μm tolerance, whereas the traditional Canny operator reaches only 17.1%, rendering it unsuitable for high-precision industrial applications. While the standard deviation (STD) of SAM-Canny (5.24 μm) was slightly higher than that of Canny (4.52 μm), this minor trade-off is acceptable given the substantial improvements in accuracy and reliability.

These results indicate that the integrated approach substantially improves measurement accuracy and robustness. The method successfully addresses the core challenges of automated frame measurement, achieving performance that meets the stringent requirements for HEPS component inspection, where both precision and consistency are critical for subsequent assembly processes.

Table 1 Comparison of Measurement Error Metrics

Method	Performance Metrics			
	Emax (μm)	Bias (μm)	STD (μm)	WTR
SAM-Canny	18.15	-6.49	5.24	92.9%
Canny	28.31	20.61	4.52	17.1%

4 Error compensation experiment

4.1 Error Compensation Experiment Dataset

During the construction of the diameter measurement dataset, all 284 images recorded the corresponding illumination intensity. The SAM3 semantic segmentation model was employed to precisely segment not only the CuCrZr frames but also other regions composed of different materials. After segmentation, the mean grayscale value of each material surface was used as the input feature, while the output was defined as the diameter error compensation coefficient, thereby constructing an experimental dataset for error compensation. Based on this dataset, the performance advantages of the

WOA-RBF neural network over conventional linear fitting methods in diameter error compensation tasks were further evaluated.

4.2 Statistical indicators

To quantitatively evaluate the performance of the diameter measurement algorithm after error compensation, five metrics are employed: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and Within-Tolerance Rate (WTR). Let D_i^{pred} denote the predicted diameter after error compensation for the i -th sample, and D_i^{true} the corresponding ground-truth diameter, with N samples in total.

The mean squared error (MSE) measures the average squared difference between predicted and true diameters, capturing overall deviation:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \left(D_i^{\text{pred}} - D_i^{\text{true}} \right)^2 \quad (28)$$

The root mean squared error (RMSE) is the square root of MSE, providing an error measure in the same units as the diameter:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(D_i^{\text{pred}} - D_i^{\text{true}} \right)^2} \quad (29)$$

The mean absolute error (MAE) represents the average magnitude of the errors, without considering their direction:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \left| D_i^{\text{pred}} - D_i^{\text{true}} \right| \quad (30)$$

The coefficient of determination (R^2) evaluates how well the predicted diameters match the variance of the ground-truth values:

$$R^2 = 1 - \frac{\sum_{i=1}^N \left(D_i^{\text{pred}} - D_i^{\text{true}} \right)^2}{\sum_{i=1}^N \left(D_i^{\text{true}} - \bar{D}^{\text{true}} \right)^2} \quad (31)$$

Finally, the Within-Tolerance Rate (WTR) is defined as the proportion of samples for which the absolute measurement error after compensation is within $10 \mu\text{m}$:

$$\text{WTR} = \frac{1}{N} \sum_{i=1}^N \mathbb{I} \left(\left| D_i^{\text{pred}} - D_i^{\text{true}} \right| \leq 10 \mu\text{m} \right) \times 100\% \quad (32)$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Here, $\bar{D}^{\text{true}}$ denotes the mean of the ground-truth diameters. Among these metrics, WTR reflects the fraction of measurements meeting high-precision requirements and serves as a practical indicator of the effectiveness of the error compensation.

4.3 Error Compensation Model Implementation Details

To comparatively analyze the influence of varying input dimensions and model complexities on the prediction of the Diameter Error Compensation Coefficient, this study constructed three predictive models: the Least Squares Method (Univariate Linear Regression), the Genetic Algorithm-Optimized Least Squares Support Vector Machine (GA-LSSVM), and the Whale Optimization Algorithm-Optimized Radial Basis Function Neural Network (WOA-RBF). The output $\mathbf{y}$ for all models was consistently the final column in the raw dataset, representing the Diameter Error Compensation Coefficient. However, a fundamental distinction lies in the selection of input features $\mathbf{X}$: the Least Squares Method and GA-LSSVM were designed as Univariate Regression Models, utilizing only the first column (Light Source Intensity) as input; conversely, the WOA-RBF was formulated as a Multivariate Regression Model, employing all columns except the last (Grayscale Values of Different Material Surfaces) as its input vector. In terms of data preprocessing, the inputs and outputs for both GA-LSSVM and WOA-RBF were scaled to the $[0, 1]$ range using MinMaxScaler. Final performance assessment for the Least Squares and GA-LSSVM models was based on a 70% training set / 30% test set split (`random.state = 42`). The Least Squares method solved the linear equation directly on the training set, requiring no optimization. For GA-LSSVM, a Genetic Algorithm (GA) was employed to optimize the penalty parameter C $[0.1, 100.0]$ and the RBF kernel parameter γ $[0.001, 1.0]$. The GA was configured with a population size of 20 and a maximum iteration count of 20, with the fitness function calculated as the Mean Squared Error (MSE) in original units on an internal 80% training / 20% validation split. For the WOA-RBF network, the Whale Optimization Algorithm (WOA) was used to optimize the network’s structural and training parameters: RBF Center Count $[10, 40]$, RBF Spread Constant σ $[0.1, 2.0]$, and Learning Rate lr $[0.001, 0.02]$. The WOA was configured with a population size of 10 and a maximum iteration count of 10, utilizing 10-fold cross-validation to determine the fitness based on the normalized MSE. Finally, the WOA-RBF model, using the optimal parameters, was trained for 200 epochs on the entire normalized dataset using the Adam optimizer and MSELoss function. The predictive performance of all models was uniformly quantified using five metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and the Percentage of Samples with Absolute Error > 5 .

4.4 Experimental results and analysis of error compensation

The performance of different error compensation methods is comprehensively evaluated in Table 2. The proposed WOA-RBF model demonstrates superior performance across all key metrics. It achieves the lowest MSE (13.72), RMSE (3.70), and MAE

(2.95), indicating the smallest overall prediction errors and highest accuracy. Furthermore, WOA-RBF attains the highest R^2 value (0.96), suggesting the best fit to the data variance and strongest explanatory power.

Most notably, the WOA-RBF model achieves the highest Within-Tolerance Rate (WTR) of 82.3%, significantly outperforming 1D Least Squares (60.3%) and GA-LSSVM (60.6%). This metric represents the proportion of samples whose absolute measurement error is within the $10\text{ }\mu\text{m}$ tolerance threshold. The substantial improvement from approximately 60% to 82.3% indicates that WOA-RBF successfully corrects measurement errors for a much larger proportion of samples, with over four-fifths now meeting the stringent precision requirement.

These results validate the effectiveness of the multidimensional grayscale feature input strategy combined with the WOA-optimized RBF neural network. By leveraging material-specific grayscale information extracted from SAM segmentation masks, the model captures the complex nonlinear relationship between illumination conditions and measurement errors more accurately than conventional approaches that rely on single illumination parameters. The superior performance of WOA-RBF across all evaluation metrics confirms its robustness and practical utility for high-precision visual measurement applications.

Table 2 Performance Comparison of Different Methods

Method	Error Metrics				
	MSE	RMSE	MAE	R^2	WTR
1D Least Squares	42.05	6.48	5.15	0.89	60.3%
GA-LSSVM	24.37	4.94	4.16	0.94	60.6%
WOA-RBF	13.72	3.70	2.95	0.96	82.3%

5 Generalization experiment

5.1 Generalization Experiment System and Procedure

To verify the robustness and generalization ability of the proposed image-based measurement algorithm and error compensation model, the generalization experiment utilized the same high-precision measurement platform and reference equipment as the training experiment. The core of the experimental platform is the image acquisition module based on a high-resolution CCD system. This module employs a GP-660V microscopic CCD camera with a resolution of 1920×1080 pixels, illuminated by a D-100 ring light source. Prior to formal measurement, the system underwent precise geometric parameter calibration using a 7×7 circular calibration target (circle diameter 0.625 mm, inter-center spacing 1.25 mm). This calibration solved for the camera’s intrinsic and extrinsic parameters and compensated for lens distortion, ensuring the geometric accuracy of the measurement results. The experiment control and image processing were executed by a computer equipped with an Intel Core i5-12400F

processor and 16 GB of memory. For the experimental subjects, the generalization experiment selected an additional 5 new CuCrZr frames. The ground-truth diameters of these frames were obtained using a KEYENCE VK-X3000 laser confocal microscope (equipped with a 20 $\times$ objective lens, spatial resolution 686.688 nm/pixel), achieving nanometer-level accuracy to serve as a reliable reference for performance evaluation. In the experimental procedure, the researchers adjusted the camera system’s light source configuration to obtain images with clear feature edges and suitable contrast, ensuring the image quality met measurement requirements. Subsequently, images of the new frames were acquired, the raw diameter was measured using the proposed image measurement algorithm, and finally, error correction was performed using the fixed WOA-RBF neural network model weights determined in the training experiment, comprehensively assessing the model’s generalization performance on unseen samples.

5.2 Statistical indicators

The primary metric used to evaluate the accuracy of both the manual measurement and the proposed algorithm measurement is the diameter absolute error, which is defined as follows:

$$E_{\text{abs},i} = D_{\text{measured},i} - D_{\text{true},i} \quad (33)$$

where $E_{\text{abs},i}$ is the absolute diameter error for the i -th sample; $D_{\text{measured},i}$ is the measured diameter obtained by the investigated method (manual measurement or the proposed algorithm); and $D_{\text{true},i}$ is the diameter ground-truth obtained using the high-precision reference instrument (e.g., the laser confocal microscope).

5.3 Generalization of experimental results and analysis

A comparative study was conducted between the conventional method for outer diameter inspection of CuCrZr frames and the image-based measurement method utilizing machine vision proposed in this paper. The conventional approach employed a Mitutoyo digital micrometer (Model: 326-251-30, TMG25MX) for manual measurement. The specific procedure was as follows: the frame was gently clamped between the anvil and the spindle of the micrometer; it was then rotated along its effective length, and multiple positions were selected for outer diameter measurement. A total of 10 sets of diameter data were collected, and the arithmetic mean of these 10 measurements was taken as the representative diameter of the frame.

The generalization capability of the proposed measurement system was rigorously evaluated on five previously unseen CuCrZr frames, with the results detailed in Table 3. The algorithm consistently maintained low measurement errors across all test samples, with absolute errors ranging from 2.4 μm to 6.3 μm . In contrast, manual measurement errors exhibited significant variability, spanning from 1.3 μm to 32.1 μm . These results clearly demonstrate that the machine vision approach provides significantly more stable and consistent measurements compared to manual operations, which are prone to human error and variability.

Table 3 Comparison of Measurement Errors(Unit: μm) for Chrome-Zirconium-Copper Frames

Measurement Error Type	Chrome-Zirconium-Copper Frame ID				
	1#	2#	3#	4#	5#
Manual Measurement Error	3.7	10.3	32.1	8.7	1.3
Algorithm Measurement Error	2.4	2.7	-4.3	6.3	4.2

6 Conclusion

This study successfully developed and validated a novel, fully automated vision-based measurement system for the high-precision diameter inspection of CuCrZr frames, a key component for the HEPS. To overcome the challenges of low manual efficiency, illumination sensitivity, and parameter-tuning dependency inherent in conventional methods, this work presents a comprehensive framework that strategically integrates zero-shot segmentation, adaptive edge detection, and data-driven error compensation. The validity and robustness of the proposed system were definitively confirmed through generalization tests on new frames. The algorithm consistently maintained measurement errors within single-digit micrometers, significantly outperforming manual measurements which exhibited errors of up to $32.1 \mu\text{m}$. This result substantiates the superior accuracy, reliability, and repeatability of the proposed automated approach.

This study presents an automated framework for micrometer-level diameter measurement of CuCrZr frames. The integration of SAM-based segmentation, adaptive edge detection, and multidimensional error compensation enables stable performance across varying illumination conditions. Although the system demonstrates strong accuracy and generalization, future work will focus on reducing the dependence on manually specified prompts and improving processing speed for real-time deployment.

Acknowledgements. This work was supported by the National Natural Science Foundation of China under Grant Nos. 12405374 and 12475330, and by the Special Program for Science and Technology Platforms of the Chinese Academy of Sciences under Grant No. 2025000148.

References

- [1] Serebrennikov, D., Clementyev, E., Semenov, A., *et al.*: Optical performance of materials for x-ray refractive optics in the energy range 8–100 kev. *Journal of Synchrotron Radiation* **23**(6), 1315–1322 (2016)
- [2] Wang, D., Zhao, W.S., Gu, L., *et al.*: A study on micro-hole machining of polycrystalline diamond by micro-electrical discharge machining. *Journal of Materials Processing Technology* **211**(1), 3–11 (2011)

- [3] Antipov, S., Baryshev, S.V., Butler, J.E., *et al.*: Single-crystal diamond refractive lens for focusing x-rays in two dimensions. *Journal of Synchrotron Radiation* **23**(1), 163–168 (2016)
- [4] Zholudev, S.I., Terentiev, S.A., Polyakov, S.N., *et al.*: Imaging by 2D parabolic diamond X-ray compound refractive lens at the laboratory source. In: *AIP Conference Proceedings*, vol. 1764, p. 020006. AIP Publishing, Melville, NY, USA (2016). International Conference on X-Ray Lasers and Coherent X-Ray Sources, Prague, Czech Republic
- [5] Polikarpov, M., Polikarpov, V., Snigireva, I., *et al.*: Diamond x-ray refractive lenses with high acceptance. *Physics Procedia* **84**, 213–220 (2016)
- [6] Medvedskaya, P., Lyatun, I., Shevyrtalov, S., *et al.*: Diamond refractive micro-lenses for full-field x-ray imaging and microscopy produced with ion beam lithography. *Optics Express* **28**(4), 4773–4785 (2020)
- [7] Celestre, R., Antipov, S., Gomez, E., *et al.*: Polished diamond x-ray lenses. *Journal of Synchrotron Radiation* **29**(3) (2022)
- [8] Wen, X., Wang, J., Zhang, G., Niu, L.: Three-dimensional morphology and size measurement of high-temperature metal components based on machine vision technology: A review. *Sensors* **21**(14), 4680 (2021) <https://doi.org/10.3390/s21144680>
- [9] Guo, J., Yang, Y., Chen, X.: Research on sub-pixel accuracy flange disk dimension measurement based on machine vision. *Signal, Image and Video Processing* **18**, 6185–6195 (2024) <https://doi.org/10.1007/s11760-024-03306-x>
- [10] Li, B.: Research on geometric dimension measurement system of shaft parts based on machine vision. *EURASIP Journal on Image and Video Processing* **2018**(1) (2018)
- [11] Fang, Y., Wang, X., Xin, Y., Luo, Y.: Sub-pixel dimensional and vision measurement method of eccentricity for annular parts. *Applied Optics* **61**, 1531–1538 (2022)
- [12] Zhang, X., Fu, F., Fan, J., Fan, H.: Design of high precision image measurement system for small workpiece. In: *Journal of Physics: Conference Series*, vol. 1865, p. 042140. IOP Publishing, Bristol, UK (2021). <https://doi.org/10.1088/1742-6596/1865/4/042140>. International Conference on Advances in Optics and Computational Sciences (ICAOCs 2021), Ottawa, Canada, 21–23 Jan. 2021
- [13] Bamerni, S.M.H., *et al.*: Recent advances on image edge detection: A comprehensive review. *Expert Systems with Applications* **214**, 119197 (2023)
- [14] Sonnleitner, M., Jeffers, J., Barnett, S.M.: Image retrodiction at low light levels.

- Optica **2**(9), 950–957 (2015) <https://doi.org/10.1364/OPTICA.2.000950>
- [15] Hynecek, J., Nishiwaki, T.: Excess noise and other important characteristics of low light level imaging using charge multiplying ccds. IEEE Transactions on Electron Devices **50**(1), 239–245 (2003) <https://doi.org/10.1109/TED.2002.806962>
 - [16] Su, X.-m., Fang, C.-g., Pan, Y.-b., Wu, W.-w., Li, Y.-p., Zhu, L.: Modeling error of visual measurement system under changing illuminance. Journal of Zhejiang University (Engineering Science) **54**(10), 1929–1935 (2020) <https://doi.org/10.3785/j.issn.1008-973X.2020.10.009>
 - [17] Kirillov, A., Mintun, E., Ravi, N., et al.: Segment Anything (2023)
 - [18] Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
 - [19] Ghosal, S., Mehrotra, R.: Orthogonal moment operators for subpixel edge detection. Pattern Recognition **26**(2), 295–306 (1993)
 - [20] Zhang, X., Fu, F., Fan, J., Fan, H.: Design of high precision image measurement system for small workpiece. Journal of Physics: Conference Series **1865**, 042140 (2021) <https://doi.org/10.1088/1742-6596/1865/4/042140> . Published under licence by IOP Publishing Ltd