



Article

One Method for Improving Overlay Accuracy Through Focus Control

Yanping Lan ¹, Jingchao Qi ^{1,*} and Mengxi Gui ²¹ Shanghai Yuwei Semiconductor Technology Co., Ltd., Shanghai 201203, China² Changxin Memory Technologies, Hefei 230601, China

* Correspondence: qjic@yuweitek.com

Abstract

Image-Based Overlay (IBO) equipment leverages optical reflection imaging principles, combined with focusing and alignment strategies to measure overlay marks. Among all measurement steps, the focal plane measurement of marks exerts the most critical impact on overlay accuracy, while the time consumed by focal plane detection directly determines the overall measurement throughput. To address the trade-off between accuracy and efficiency in advanced process nodes, this paper proposes an integrated optimization strategy encompassing optical hardware design and software algorithms. The hardware solution adopts a dual-wavelength, dual-detector architecture: optimal imaging wavelengths are selected independently for the previous-layer and current-layer marks, ensuring each layer achieves ideal imaging conditions without mutual interference. The software strategy employs a deep learning framework to simultaneously predict and adjust the horizontal position (alignment) and vertical defocus number of measured marks in real time with high precision, thereby securing the optimal imaging posture. By synergizing hardware-based optimal imaging conditions and software-based posture adjustment, this method effectively mitigates the impact of background noise and system aberrations, ultimately improving both the accuracy and efficiency of overlay measurement.

Keywords: IBO; overlay; dual-wavelength optical; focal plane measurement; mark alignment

1. Introduction

Overlay accuracy, alongside resolution and throughput, stands as one of the three core performance indicators of lithography technology [1,2]. It directly dictates the electrical connectivity of multi-layer circuits and the final yield of semiconductor chips. With the advancement of process nodes into the 14 nm, 7 nm, and even 3 nm regimes, the critical dimension (CD) of devices has shrunk to single-digit nanometers, requiring overlay accuracy to be controlled below 2 nm (3σ) [3]. This places unprecedented demands on overlay metrology systems, which primarily consist of two technical routes: Image-Based Overlay (IBO) and Diffraction-Based Overlay (DBO) [4].

IBO equipment leverages optical reflection imaging principles to directly capture mark images, offering intuitive interpretation and robustness against certain process variations. In contrast, DBO relies on diffracted signals from periodic grating marks, extracting overlay information through spectral analysis. While DBO typically achieves higher precision and repeatability for certain film stacks due to its sensitivity to phase information, it often requires dedicated grating targets and is more susceptible to process-induced asymmetries and thin-film interference effects. Moreover, DBO systems may struggle with complex



Academic Editors: Sikun Li and Yayi Wei

Received: 30 October 2025

Revised: 15 December 2025

Accepted: 19 December 2025

Published: 2 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

multi-layer stacks where multiple materials induce spectral crosstalk, limiting material versatility. IBO, with its direct imaging approach, offers greater flexibility in mark design and compatibility with diverse film systems, though traditionally at the cost of lower throughput and susceptibility to focus-induced errors.

The proposed method addresses key limitations of traditional IBO—namely, wavelength compatibility, throughput, and noise sensitivity—through an integrated hardware and software co-design. Compared to DBO, our approach retains IBO’s material versatility and adaptability to existing Box-in-Box marks while significantly closing the gap in precision and repeatability. By enabling dual-wavelength simultaneous imaging and deep learning-based joint alignment-focusing, our solution achieves sub-nanometer focal plane control and high throughput, making it competitive with DBO in advanced nodes, particularly in environments with complex film stacks and high process variability.

1.1. Related Work

Recent studies have attempted to improve IBO accuracy through optical optimization or machine learning. Multi-color per-layer imaging approaches have been proposed to minimize crosstalk in advanced film stacks, while ML-based residual reduction methods focus on TIS modeling [5]. However, these methods still rely on sequential focusing and alignment and typically use a single camera.

Deep learning has been applied to mark detection [6] and classification [7], but joint alignment–focus prediction remains unexplored. Moreover, prior multi-wavelength methods rely on sequential illumination rather than true parallel acquisition. The proposed method introduces simultaneous dual-wavelength acquisition with integrated deep-learning-based optimization, offering a new solution pathway.

1.2. Principles of IBO Measurement

IBO equipment operates on the principle of optical reflection: an optical system projects overlay marks (consisting of previous-layer and current-layer patterns) onto an image sensor (camera). Post-imaging, specialized software analyzes the centroid positions of the two layers of marks; the positional offset between these centroids is defined as the overlay error (Misregistration). Common overlay mark designs include Box-in-Box, Bar-in-Bar, and Advanced Imaging Metrology (AIM) marks [8], with Box-in-Box being the most widely used in mass production due to its simplicity and robustness.

A typical Box-in-Box overlay mark is shown in Figure 1, consisting of two concentric rectangular frames. The gray outer frame represents the previous-layer mark, with its centroid at coordinates (x_p, y_p) ; the black inner frame represents the current-layer mark, with its centroid at (x_c, y_c) . Each frame comprises four orthogonal bars: Top, Bottom, Left, and Right, all with identical length ($20\ \mu\text{m}$) and width ($1\sim 2\ \mu\text{m}$) in this example. For horizontal (x-direction) overlay measurement, the Left and Right bars are used (y-direction serves as the non-measurement axis); for vertical (y-direction) measurement, the Top and Bottom bars are used (x-direction serves as the non-measurement axis). The overlay error is calculated as: Misregistration $X = x_c - x_p$ (horizontal offset) and Misregistration $Y = y_c - y_p$ (vertical offset).

A flowchart illustrating the sequential focusing and alignment process in traditional IBO systems is illustrated in Figure 2. In Figure 2a, the x-axis denotes the focus position of the objective lens, scanned in discrete increments, while the y-axis represents the computed contrast of the alignment mark image. The nested square patterns illustrate the alignment mark at different focus positions, showing increasing sharpness as the lens approaches the optimal focal plane and decreasing sharpness as it moves away. The dotted curve plots contrast against focus position, forming a distinct peak; the apex of this peak identifies the

optimal focal plane, where mark edge definition is maximized to enable precise overlay measurement. In Figure 2b, colors in the template and test images represent different alignment mark edges, used to determine horizontal position (x , y) via template matching.

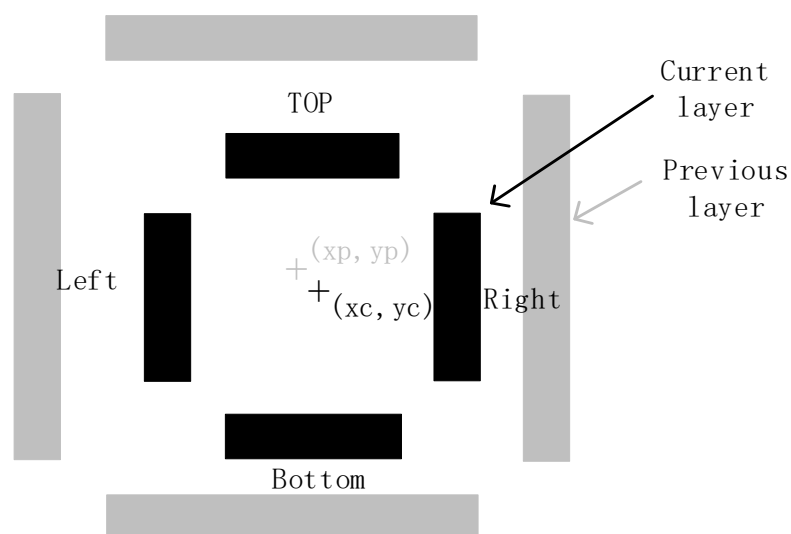


Figure 1. Bar-In-Bar overlay mark.

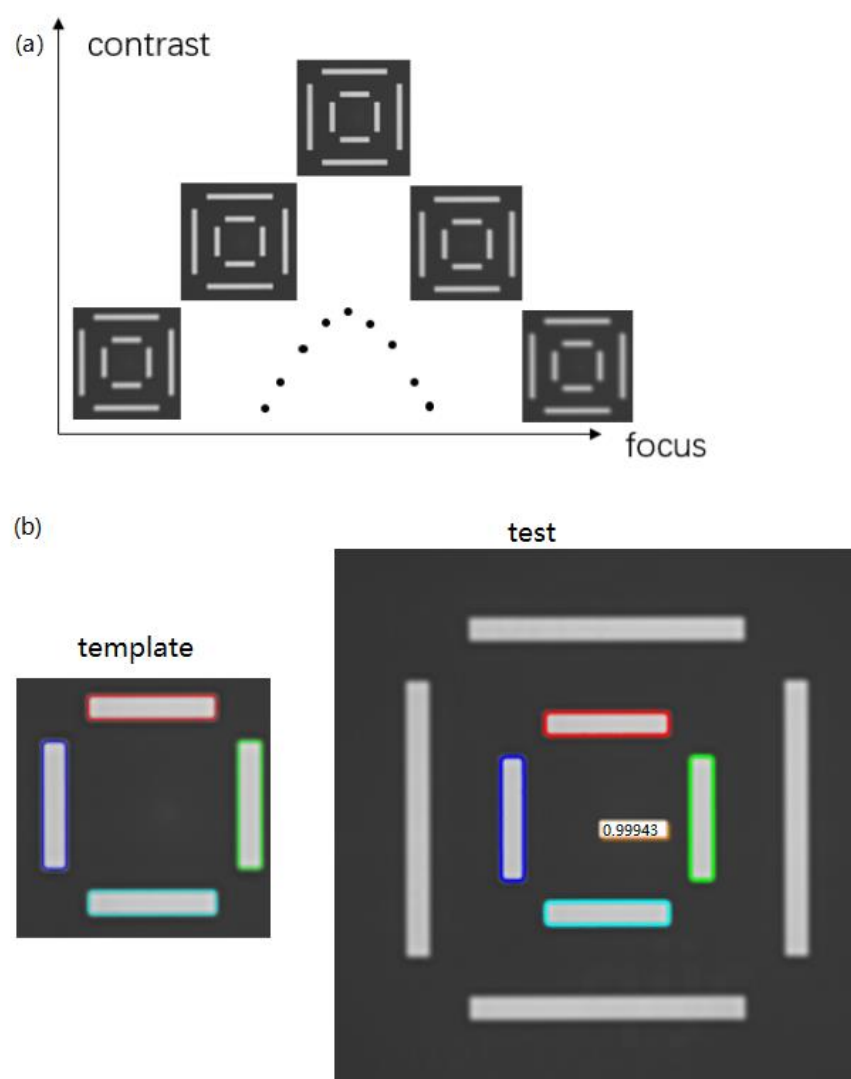


Figure 2. Contrast Calculated Focal Plane and Template Matching Alignment.

The workflow starts with “Rough Focusing” (step 1), where the objective lens moves in 100 nm increments to capture a series of low-contrast images. Next, “Template Matching” (step 2) compares these images to a pre-stored mark template to determine the horizontal position (x, y) of the mark and center it in the FOV (reducing symmetric aberration effects). “Fine Focusing” (step 3) then uses a contrast calculation algorithm (e.g., Laplacian variance) to find the focal plane with the highest edge sharpness (defocus amount = 0). Finally, “Image Capture” (step 4) acquires the high-quality image for overlay analysis. The total time for steps 1–3 is ~50 ms, with fine focusing accounting for 60% of this duration.

1.3. Limitations of Traditional IBO Schemes

Traditional IBO systems adopt a single optical path, single-wavelength, and single-camera configuration, paired with a high-numerical-aperture ($NA \geq 0.7$) achromatic objective lens to reduce chromatic aberration. However, this design faces three key challenges in advanced processes:

1. **Wavelength Compatibility Issues:** For complex film stacks (e.g., SiN hard masks, EUV photoresists), a single wavelength cannot simultaneously ensure clear imaging of both previous-layer and current-layer marks. For instance, visible light (405–633 nm) is blocked by SiN films (>200 nm thick), while near-infrared (NIR) light (900–1100 nm) struggles to resolve fine current-layer marks due to lower diffraction-limited resolution [9].
2. **Low Throughput from Sequential Focusing/Alignment:** To ensure optimal imaging, traditional schemes require sequential operations. This sequential workflow takes about 50 ms per mark, limiting the throughput to <30 wafers per hour for high-density sampling (200 marks per wafer) [2].
3. **Background Noise Interference:** During image acquisition, the sensor captures not only the signal from the mark but also background signals (e.g., scattered light from the wafer surface, reflections from underlayers). This reduces the signal-to-noise ratio (SNR) of mark edges, introducing measurement residuals of up to 0.3 nm [8].

To address these limitations, this paper proposes a dual-wavelength, dual-camera hardware architecture combined with a deep learning-based joint alignment-focusing algorithm. This integrated solution enables parallel imaging of previous and current layers, while simultaneously optimizing mark position and focal plane—thus improving both accuracy and throughput.

2. Dual-Wavelength Optical Scheme

The selection of the light source is pivotal to IBO performance, as wavelength directly impacts mark contrast, film penetration, and measurement repeatability. The evolution of IBO light sources and their corresponding performance metrics are summarized in Table 1. Over time, IBO light sources have evolved from white halogen lamps (broad spectrum, low intensity) to laser-driven light sources (LDLS) with color filters, and finally to wavelength tuners (continuous wavelength adjustment), as shown in Table 1. This evolution reflects the growing demand for wavelength flexibility in advanced processes.

Table 1. Light Source Evolution and Corresponding TMU.

Process Node	Light Source	TMU (nm)	Key Limitation
>28 nm	White Halogen Lamp	0.75	Poor contrast for thin films; high noise
28~14 nm	LDLS + Color Filters	0.43	Limited wavelength options; cross-layer interference
<14 nm	LDLS + Wavelength Tuner	0.31	Single wavelength cannot optimize both layers

2.1. Evolution of IBO Light Sources

For nodes < 14 nm, the complexity of film stacks (e.g., multi-layer SiN/SiO₂, metal gates) makes it impossible to find a single wavelength that satisfies both “previous-layer penetration” and “current-layer resolution” [9]. As shown in Figure 3, the previous-layer mark (under a 300 nm SiN film) achieves maximum contrast at 940 nm (NIR), while the current-layer mark (EUV photoresist) performs best at 405 nm (deep ultraviolet, DUV). A single wavelength (e.g., 633 nm) results in contrast values < 20 for both layers, leading to centroid extraction errors.

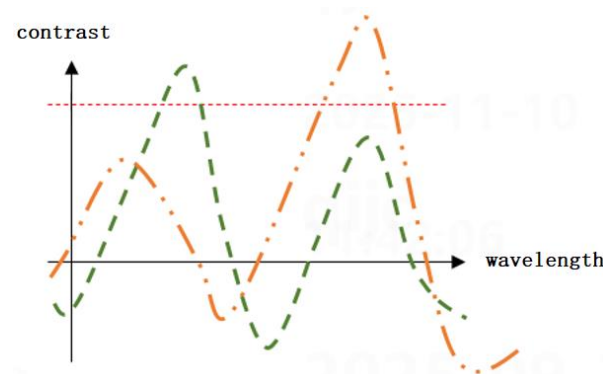


Figure 3. Contrast and Optimal Wavelengths for Two Layers.

Table 1 illustrates the evolution of overlay metrology light source technologies and their corresponding TMU performance across different semiconductor process nodes. TMU (Total Measurement Uncertainty) is a key indicator evaluating the precision of overlay metrology systems, with smaller values representing higher measurement accuracy, where

$$TMU = \sqrt{TIS_{3\delta}^2 + TIS_{Mean}^2 + Precision^2}.$$

Figure 3 shows the contrast (y-axis) vs. wavelength (x-axis) curves for previous-layer and current-layer marks. The orange dashed line represents the previous-layer mark (under a 300 nm SiN hard mask): contrast increases with wavelength, peaking at 940 nm due to enhanced penetration of NIR light through SiN. The green curve represents the current-layer mark (100 nm EUV photoresist): contrast peaks at 405 nm because DUV light provides higher diffraction-limited resolution for fine features. The red dashed horizontal line serves as a KPI threshold. Wavelength ranges where the curves exceed this threshold are deemed optimal, as the KPI meets or surpasses the required performance standard in these intervals. This validates the need for dual wavelengths to optimize each layer independently.

2.2. Dual-Wavelength, Dual-Camera Hardware Design

The dual-wavelength optical scheme retains the core reflection imaging principle of traditional IBO but modifies the hardware to enable parallel, layer-specific imaging. The key components include: (1) two independent LDLS light sources with wavelength tuners; (2) a dichroic prism (beam splitter); (3) two high-resolution cameras; and (4) a shared objective lens (to ensure coaxiality of the two optical paths).

This optical scheme is a high-precision dual-wavelength, dual-camera optical system for semiconductor Overlay measurement, achieving high-precision alignment and measurement of wafers through multi-module collaboration. Figure 4 shows the Dual-Wavelength Optical Path. It mainly consists of three parts: light source module, illumination and optical path module, detection module.

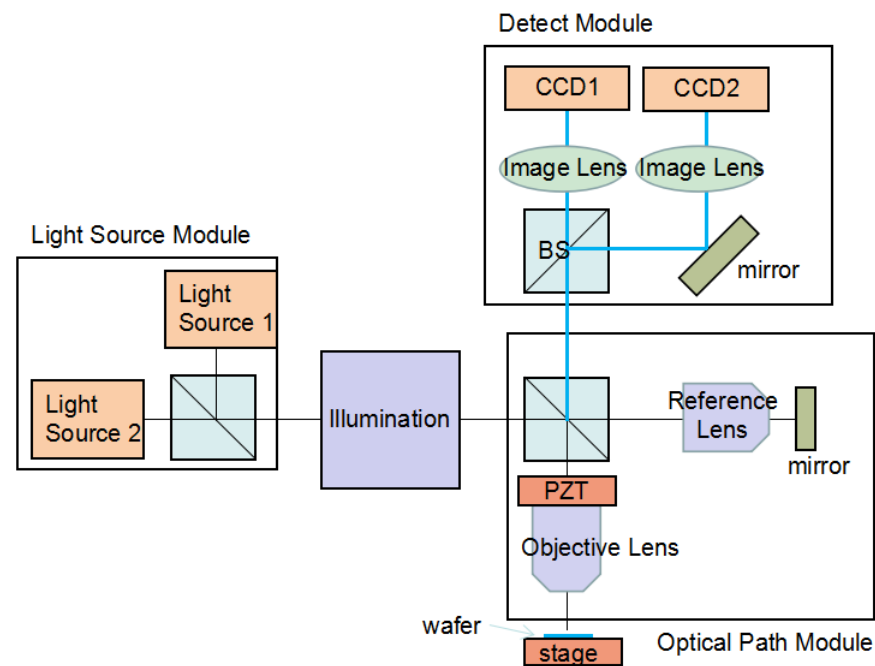


Figure 4. Dual-Wavelength Optical Path.

1. **Light Source Module:** The light source module (Light Source Module) is equipped with two light sources (Light Source 1, Light Source 2) that can output light of two different wavelength bands (e.g., visible light & near-infrared, deep ultraviolet & infrared) to achieve “dual-wavelength” division of labor. One wavelength is used to penetrate multi-layer structures such as hard masks and photoresists on the wafer surface to identify lower-layer marks, while the other wavelength is used to enhance the edge contrast of upper-layer patterns for precise positioning of upper-layer marks. The two light sources are combined into a single beam via a beam-combining element before entering the illumination system.
2. **Illumination and Optical Path Module:** The illumination unit (Illumination) homogenizes and collimates the combined light to ensure light field uniformity. The light is split into a measurement optical path and a reference optical path. In the measurement optical path, the light is focused onto the wafer surface via an objective lens (Objective Lens). The wafer is placed on a high-precision wafer stage that enables nanometer-level motion control. After reflection by the wafer, the light carries structural information of the measured pattern and returns. In the reference optical path, the light forms reference light via a reference lens (Reference Lens) and a mirror (Mirror), which is used for subsequent interference or image comparison to eliminate systematic errors. Among them, a piezoelectric ceramic (PZT) can drive optical components to achieve nanometer-level displacement for precise focusing or phase modulation, further improving measurement accuracy.
3. **Detection Module:** The detection module (Detector Module) adopts a dual-camera (CCD1, CCD2) architecture. A beam splitter (BS) separates the dual-wavelength light carrying measured information, which is then focused via imaging lenses (Image Lens) and synchronously captured by two CCD cameras (i.e., the “Double Grab” function). The synchronization of the two cameras ensures that the upper and lower layer marks are imaged at the same time and in the same FOV, providing a foundation for accurate registration of subsequent Overlay algorithms. In addition, beam-splitting elements can adjust the beam-splitting ratio or imaging magnification of the optical path to adapt to the accuracy requirements of different measurement scenarios.

The advantages of this design:

1. Independent Wavelength Optimization: Each layer uses its optimal wavelength (e.g., 940 nm for previous-layer SiN penetration, 405 nm for current-layer resolution), maximizing mark contrast.
2. Parallel Imaging: Both layers are imaged simultaneously (Double Grab, DG) instead of sequentially, reducing per-mark measurement time.
3. Reduced Cross-Layer Interference: Since each camera only captures one wavelength, there is no signal crosstalk between previous and current layers—critical for thin-film stacks with high light scattering.

3. Alignment and Focusing Model Based on Deep Learning

While the dual-wavelength hardware improves imaging quality, optimal overlay accuracy still requires precise alignment (horizontal position) and focusing (vertical focal plane) of the mark. Traditional sequential methods (template matching + contrast calculation) are slow and prone to noise. To address this, we propose a multi-task deep learning model that simultaneously predicts the mark's horizontal bounding box (alignment) and optimal focal plane (defocus amount) from a single image.

3.1. Model Inputs and Outputs

The model takes two types of inputs to account for both image information and system variability:

1. Image Input: A grayscale image of the overlay mark (resolution: 512×512 pixels; FOV: $10 \mu\text{m} \times 10 \mu\text{m}$), captured under either DUV or NIR light.
2. Optical Parameter Input: A K-dimensional vector of system parameters ($K = 5$ in this study) that affect imaging, including: wavelength (nm), light source intensity (mW), objective lens NA, camera exposure time (μs), and wafer film thickness (nm).

The model outputs two key results for overlay measurement:

3. Alignment Output: A bounding box (x, y, w, h) defining the mark's position in the FOV (x, y = top-left corner coordinates; w, h = width/height of the box).
4. Focus Output: The optimal focal plane position (z , in nm), defined as the vertical distance between the objective lens and the mark surface that maximizes edge contrast.

3.2. Model Architecture

The architecture adopts an end-to-end design, integrating image feature extraction, optical parameter encoding, feature fusion, and multi-task prediction (Figure 5).

Figure 5 Description: The model architecture consists of five core modules: (1) Backbone Network (Image Feature Extraction); (2) Optical Parameter Encoder; (3) Feature Fusion; (4) Multi-Task Head (Detection + Depth); and (5) Loss Function. Detailed breakdown:

1. Backbone Network (Image Feature Extraction): ResNet-101 is used to extract multi-scale image features, outputting three feature maps (C3: $64 \times 64 \times 256$; C4: $32 \times 32 \times 512$; C5: $16 \times 16 \times 1024$) after $8\times$, $16\times$, and $32\times$ downsampling. A Feature Pyramid Network (FPN) then generates P3–P7 feature maps (P3: $64 \times 64 \times 256$; P7: $2 \times 2 \times 256$) to capture both fine-grained (P3) and global (P7) features.
2. Optical Parameter Encoder: A 2-layer fully connected (FC) network processes the 5-dimensional optical parameters, outputting a 256-dimensional feature vector (h_{opt}) via ReLU activation. Optical parameters are encoded into feature vectors compatible with image features.

3. Feature Fusion (Image Features and Optical Parameters): h_opt is spatially expanded to match the size of P3 ($64 \times 64 \times 256$) using bilinear interpolation, then fused with P3–P5 via element-wise addition (residual connection) to inject system parameter information into image features.
4. Multi Task Head contains detection head and depth head. Detection Head (Alignment) Follows the Faster R-CNN framework: a Region Proposal Network (RPN) generates anchor boxes on P3–P5; ROI Align extracts $7 \times 7 \times 256$ features for each anchor; two FC layers predict class confidence (mark vs. background) and bounding box offsets. Depth Head (Focus) Uses a Squeeze-and-Excitation (SE) Block to weight ROI features (enhancing edge-related regions), followed by three FC layers to regress the focal plane z-value.
5. Loss Function: Total loss = $\lambda_1 \times \text{Detection Loss} + \lambda_2 \times \text{Depth Loss}$. $\lambda_1 = 1.0$, $\lambda_2 = 0.5$; hyper parameters tuned via cross-validation. The values of $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$ in the loss function are based on 5-fold cross-validation: four combinations of $\lambda_2 = 0.3, 0.5, 0.7$, and 1.0 were compared (with λ_1 fixed at 1.0), and it was found that when $\lambda_2 = 0.5$, the comprehensive error of alignment precision ($\text{IoU} \pm 0.3$ pixel) and focusing accuracy (± 8 nm) is minimized. Sensitivity analysis shows that when $\lambda_2 > 0.7$, the weight of focusing loss is too high, leading to an increase in alignment bounding box offset (IoU drops to ± 0.5 pixel); when $\lambda_2 < 0.3$, the weight of alignment loss is too high, resulting in an increase in focusing accuracy is ± 12 nm, neither of which can meet the requirements of advanced processes.

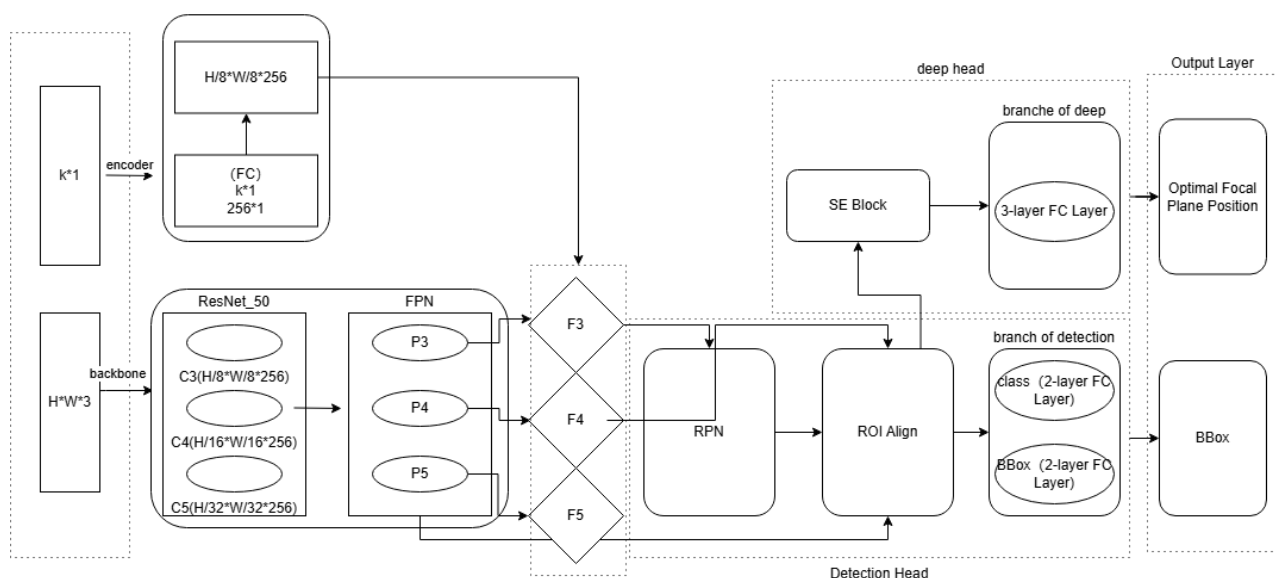


Figure 5. Schematic Diagram of the Deep Learning Model Structure.

3.3. Loss Function Definition

To balance alignment accuracy and focal plane precision, the model uses a combined loss function. The detection loss includes classification loss (Focal Loss, to address class imbalance between mark and background) and bounding box regression loss (GIoU Loss, to improve localization accuracy).

Detection loss:

$$L_{\text{det}} = L_{\text{cls}} + L_{\text{box}}$$

$$L_{\text{cls}} = -\frac{1}{N_{\text{pos}}} \sum_{i=1}^{N_{\text{pos}}} \alpha (1 - \hat{p}_i)^{\gamma \log(\hat{p}_i)} - \frac{1}{N_{\text{neg}}} \sum_{j=1}^{N_{\text{neg}}} (1 - \alpha) (\hat{p}_j)^{\gamma \log(1 - \hat{p}_j)} \quad (\text{Focal Loss})$$

$$L_{\text{box}} = 1 - \text{GIoU}(\hat{b}, b)$$

Focal plane regression loss: The focal plane regression uses L1 Loss (robust to outliers) to minimize the difference between predicted $\hat{z}$ and ground-truth z focal plane positions.

$$L_{\text{depth}} = |\hat{z} - z|$$

Total loss:

$$L_{\text{total}} = \lambda_1 L_{\text{det}} + \lambda_2 L_{\text{depth}} \quad (\lambda_1, \lambda_2 \text{ are hyperparameters})$$

3.4. Model Training and Validation

The model was trained on a dataset of 10,000 overlay mark images captured from a 7 nm DRAM wafer. Each sample included: (1) a 512×512 grayscale image; (2) 5-dimensional optical parameters; (3) ground-truth bounding box (labeled manually); (4) ground-truth focal plane (measured via high-precision interferometry). The dataset covered 50 different film stacks and 10 wavelength settings (405–940 nm).

Training was performed on an NVIDIA A100 GPU for 100 epochs, with a batch size of 32 and Adam optimizer (learning rate = 1×10^{-4}). Validation on a holdout set (2000 samples) showed the following:

Alignment accuracy: Bounding box IoU (Intersection over Union) with ground truth is ± 0.3 pixel.

Focus accuracy: Mean absolute error (MAE) between predicted and ground-truth is ± 8 nm.

4. Experimental Results and Discussion

To validate the proposed method, experiments were conducted on a 7 nm DRAM wafer (200 overlay marks per wafer) using two setups: (1) Traditional IBO (single wavelength: 633 nm; sequential alignment/focusing); (2) Proposed IBO (dual wavelength: 405 nm + 940 nm; deep learning-based alignment/focusing). Key performance metrics—contrast, alignment/focus time, and focal plane repeatability—were compared.

4.1. Imaging Contrast Improvement

Contrast is the maximum value of the gradient signal. SNR (Signal-to-Noise Ratio) is calculated as the ratio of edge signal power to background noise power. The dual-wavelength scheme nearly doubled the contrast for both layers, attributed to: (1) 940 nm NIR light's enhanced penetration through SiN (previous layer); (2) 405 nm DUV light's high resolution for current-layer photoresist marks. Higher contrast reduces centroid extraction error—critical for sub-2 nm overlay accuracy.

The quantitative comparison of contrast performance between the two schemes is presented in Table 2. This table compares the performance of two schemes (traditional single-wavelength vs. proposed dual-wavelength) in terms of mark contrast and SNR. The dual-wavelength scheme also outperforms the traditional one (38.7 ± 2.3 vs. 15.2 ± 1.5), indicating better signal quality at mark edges. In summary, the dual-wavelength scheme significantly enhances both contrast (for both layers) and SNR compared to the traditional single-wavelength approach.

Table 2. Comparison of Mark Contrast Between Single-Wavelength and Dual-Wavelength Schemes.

Scheme	Previous-Layer Contrast (Outer Frame)	Current-Layer Contrast (Inner Frame)	SNR of Mark Edges
Traditional (633 nm)	18.3 ± 2.1	12.5 ± 1.8	15.2 ± 1.5
Proposed (Dual-WL)	35.3 ± 2.4	41.2 ± 2.7	38.7 ± 2.3

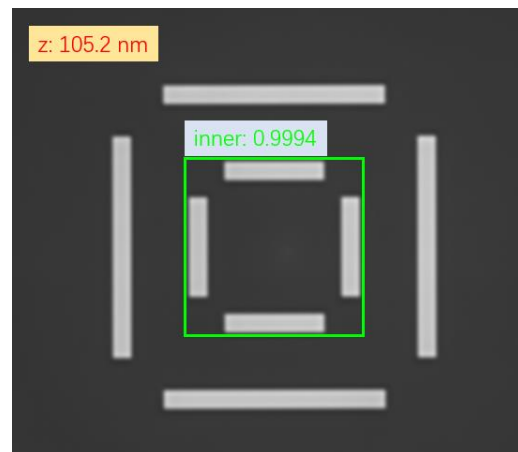
4.2. Throughput and Repeatability Improvement

The throughput and repeatability metrics of the proposed method versus the traditional scheme are presented in Table 3. The traditional scheme requires 50 ± 3 ms, while the proposed scheme reduces this to 12 ± 1 ms, achieving a 76% improvement in speed. The traditional scheme has a repeatability of 23 ± 2 nm, whereas the proposed scheme achieves 11 ± 1 nm, delivering a 52% improvement in positioning accuracy.

Table 3. Comparison of Throughput and Focal Plane Repeatability.

Metric	Traditional Scheme	Proposed Scheme	Improvement Percentage
Alignment + Focus Time per Mark	50 ± 3 ms	12 ± 1 ms	76%
Focal Plane Repeatability (3σ)	23 ± 2 nm	11 ± 1 nm	52%

A sample output from the proposed deep learning model is illustrated in Figure 6. The left panel shows the input image. The green box is the model-predicted bounding box (IoU with ground truth = 0.9994); the red text shows the predicted focal plane $z = 105.2$ nm (ground truth = 100.0 nm, error = 5.2 nm). The right panel shows the focal plane repeatability: 3σ of 100 measurements on the same mark = 8.0 nm for the proposed model, compared to 12 nm for the traditional contrast method. The reduced scatter of the proposed model (blue dots) indicates better resistance to background noise.

**Figure 6.** Focal Plane Prediction and Mark Position Prediction.

5. Conclusions

This paper presents an integrated solution for improving IBO overlay accuracy, combining a dual-wavelength, dual-camera hardware design and a multi-task deep learning algorithm. The dual-wavelength architecture enables independent optimization of imaging conditions for previous-layer and current-layer marks, doubling mark contrast and SNR compared to traditional single-wavelength schemes. The deep learning model simultaneously predicts mark alignment and focal plane from a single image, reducing alignment/focus time from 50 ms to 12 ms (76% improvement) and improving focal plane

repeatability by 52%. Experimental results on a 7 nm DRAM wafer show the proposed scheme achieves meeting the accuracy requirements.

Future work will focus on the following: (1) expanding the model to support EUV wavelengths for next-generation lithography; (2) integrating real-time film thickness estimation into the deep learning framework to further reduce calibration time; (3) miniaturizing the dual-wavelength optical system for in situ overlay measurement (embedded in lithography scanners).

Author Contributions: Conceptualization, Y.L. and J.Q. and M.G.; Validation, Y.L.; Formal analysis, Y.L.; Investigation, Y.L. and J.Q. and M.G.; Writing—original draft, Y.L.; Writing—review & editing, Y.L. and J.Q.; Visualization, Y.L. and J.Q.; Supervision, J.Q. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Yanping Lan and Jingchao Qi were employed by the Shanghai Yuwei Semiconductor Technology Co., Ltd. Author Mengxi Gui was employed by the CXMT. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lavitt, E.; Manassen, A.; Simon, Y.; Sanko, D.; Safrani, A. System and Method for Determining Target Feature Focus in Image-Based Overlay Metrology. U.S. Patent US11921825B2, 5 March 2024.
2. Li, W.H.; Wei, S.M. Automatic Recipe Optimization for Overlay Metrology Systems. U.S. Patent US20210025695, 28 January 2021.
3. Lee, H.; Han, S.; Woo, J.; Lee, D.; Song, C.; Heo, H.; Brinster, I.; Choi, D.; Robinson, J.C. High-Volume Manufacturing Device Overlay Process Control. In *Metrology, Inspection, and Process Control for Microlithography XXXI*; SPIE: Bellingham, WA, USA, 2017; p. 1014512.
4. International Technology Roadmap for Semiconductors (ITRS). 2023 Edition: Advanced Lithography. Available online: <https://www.semiconductors.org/wp-content/uploads/2018/08/ERD2003.pdf> (accessed on 15 December 2025).
5. Katz, S.; Ophir, B.; Shusterman, U.; Golotsvan, A.; Yerushalmi, L.; Megged, E.; Grauer, Y.; Zhang, J.; Shih, A.; Wei, S.-M.; et al. Machine Learning for Modeled-TIS and Overlay Reduction. In *Metrology, Inspection, and Process Control for Microlithography XXXIV*; SPIE: Bellingham, WA, USA, 2020; p. 113252C.
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
7. Fu, H.; Gong, M.; Wang, C.; Batmanghelich, K.; Tao, D. Deep Ordinal Regression Network for Monocular Depth Estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2002–2011.
8. Klein, D.; Amit, E.; Cohen, G.; Amir, N.; Har-Zvi, M.; Huang, C.-C.K.; Karur-Shanmugam, R.; Pierson, B.; Kato, C.; Kurita, H. Quality Metric for Accurate Overlay Control in <20 nm Nodes. In *Metrology, Inspection, and Process Control for Microlithography XXVII*; SPIE: Bellingham, WA, USA, 2013; p. 86811F.
9. Katz, S.; Park, S.W.; You, J.; Kim, H.; Lee, H.G.; Kim, J.; Lee, D.; Yeon, H.; Lee, J.-S.; Lee, S.-H.; et al. Residuals Reduction in Imaging-Based Overlay Using Color per Layer. In *Metrology, Inspection, and Process Control XXXVII*; SPIE: Bellingham, WA, USA, 2023; p. 124960G.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.