Results in Engineering 30 (2026) 111382

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

Multi-domain transformer IQA algorithm for AOI optical scheme evaluation ⋆

Zikang Zeng [a,1], Yiru Wang [a,1], Jinghan Wang [a], Tianwu Lei [a], Mingwei Feng [a], Ningmu Zou [a,b,*]

[a] School of Integrated Circuits, Nanjing University, 1520 Taihu Avenue, Suzhou, 215000, Jiangsu Province, China
[b] Interdisciplinary Research Center for Future Intelligent Chips (Chip-X), Nanjing University, 1520 Taihu Avenue, Suzhou, 215000, Jiangsu Province, China

A R T I C L E I N F O

Keywords:
Automatic optical inspection
Optical configuration
Defect exposure quality
WAFER-OS100 dataset

A B S T R A C T

Automatic Optical Inspection (AOI) systems require carefully designed optical configurations to ensure adequate defect visibility during image acquisition. In industrial practice, engineers typically compare images captured under different optical settings and manually select the configuration that best exposes defects. However, this process is inherently subjective and lacks quantitative modeling, making systematic optimization of optical setups challenging. In this work, we formulate AOI optical scheme selection as a supervised learning problem by modeling engineer-annotated defect exposure scores. To support this study, we construct the WAFER-OS100 dataset, which contains wafer images acquired under 100 distinct optical configurations. To effectively capture subtle defect characteristics in complex industrial environments, we propose a task-oriented learning framework that integrates frequency-aware image representations with explicit optical parameter encoding within a transformer-based architecture. By jointly modeling image content and optical parameters, the proposed method learns the relationship between optical configurations and defect exposure quality, enabling data-driven evaluation and ranking of imaging schemes. Experimental results demonstrate a strong correlation between the predicted scores and engineer annotations, indicating that the proposed framework can effectively approximate human evaluation and provide practical support for AOI optical scheme evaluation and ranking.

1. Introduction

In recent years, with the rapid growth of the manufacturing industry's demand for high-quality products, efficient and accurate quality inspection methods have become particularly important. Traditional manual inspection approaches are not only time-consuming and labor-intensive but also prone to human errors, making them increasingly inadequate to meet the stringent requirements of modern manufacturing processes. Consequently, integrating advanced technologies into manufacturing to enhance production efficiency and product quality has emerged as an inevitable trend in industrial development. As a vision-based, non-contact automated quality inspection technology, Automatic Optical Inspection (AOI) has been widely adopted for quality assessment of production line outputs [1]. For this purpose, specialized AOI test equipment has been developed and widely used on production lines to detect and identify various defects in real time [2,3], ensuring the consistency and reliability of production product quality.

AOI detection relies fundamentally on optical imaging, where the selection of the optical scheme plays a crucial role in inspection precision. Factors such as the color, angle, wavelength, and intensity of the light source can significantly affect the visibility of surface defects [4–10]. In practical industrial workflows, engineers typically evaluate multiple optical configurations by visually comparing captured images and selecting the scheme that exposes defects most clearly. However, manually designed optical schemes often fail to account for all possible anomalies that a workpiece may exhibit, and the evaluation process is tedious and prone to human bias. The lack of quantitative evaluation criteria makes systematic optimization of optical schemes challenging, particularly in semiconductor manufacturing scenarios where defects are subtle and backgrounds are complex.

⋆ This work was supported by National Natural Science Foundation of China (NSFC) under Grant (62341408), Fundamental Research Funds for the Central Universities (0213–14380265,2024300421).

* Corresponding author.
E-mail addresses: zikangzeng@smail.nju.edu.cn (Z. Zeng), yiruwang@smail.nju.edu.cn (Y. Wang), nzou@nju.edu.cn (N. Zou).

[1] Zikang Zeng and Yiru Wang contributed equally to this work.

https://doi.org/10.1016/j.rineng.2026.111382
Received 15 March 2026; Received in revised form 19 May 2026; Accepted 3 June 2026
Available online 6 June 2026

To support quantitative evaluation, Image Quality Assessment (IQA) methods [11,12] have been considered as potential tools for optical scheme assessment in AOI systems. IQA methods aim to predict perceptual image quality scores using computational models, ensuring consistency with the Mean Opinion Score (MOS) provided by human evaluators. Based on the availability of reference images, IQA methods can be categorized into Full-Reference (FR-IQA) and No-Reference (NR-IQA) approaches [13–16]. FR-IQA requires a reference image for comparison during evaluation, measuring perceptual differences between distorted and reference images. However, in AOI systems, it is difficult to define an ideal optical scheme image as a reference, since the objective is precisely to determine which configuration provides the best defect exposure. In contrast, NR-IQA does not rely on reference images but directly estimates quality scores based on the image itself, making it more suitable for practical AOI scenarios [14,16].

Nevertheless, existing NR-IQA methods are primarily designed for natural image quality evaluation and mainly rely on spatial-domain features such as texture and edge statistics [12,17]. These approaches are effective for assessing global perceptual distortions, but when dealing with small defects or complex industrial backgrounds, especially in AOI applications, they often overlook subtle defect details [1,18,19]. In particular, for tiny defects embedded in structured wafer surfaces, spatial-domain representations may fail to capture discriminative fine-grained information, thereby limiting the reliability of quality assessment for optical scheme selection. Meanwhile, most existing IQA datasets consist of natural scene images, such as LIVE [20], CSIQ [21], and KonIQ-10k [22], which do not reflect the characteristics of industrial defect images captured under varying optical configurations. As a result, existing datasets cannot effectively support training models tailored to AOI optical evaluation tasks.

To address these limitations, this paper focuses on quantitative modeling of defect visibility for optical scheme selection in AOI systems. We construct a dedicated wafer image dataset, WAFER-OS100, specifically designed for semiconductor inspection scenarios. The dataset contains images captured under 100 different optical schemes, including visible and UV illumination conditions. A robotic arm was used to adjust the angle and direction of the light source relative to the wafer, enabling multi-angle and multi-directional acquisition. In total, 91 wafers were collected, with 100 images captured under each optical scheme, resulting in 9100 images. For each image, experienced engineers evaluated defect visibility and assigned numerical scores based on the clarity of defect exposure. We define this manually assigned defect visibility score as the Defect Exposure Index (DEI), which is used as the supervision target for model training. These annotations provide structured supervision signals for modeling the relationship between optical configuration and perceived defect visibility Fig. 1.

Building upon this dataset, we develop a task-oriented framework, referred to as MDTNet, which integrates spatial-domain features, frequency-domain representations, and optical scheme information to support AOI optical evaluation. MDTNet employs WideResNet50 [23] to extract multi-scale features. The extracted features are further decomposed into low-frequency and high-frequency components through wavelet transform, capturing both global structural information and fine-grained defect textures. The frequency-domain components are fused with learned weights to form comprehensive image representations. In addition, optical scheme information (OSI) is encoded and integrated with image features to model the interaction between imaging conditions and defect appearance. Through this multi-domain fusion strategy, the proposed framework enhances sensitivity to subtle defect characteristics while maintaining robustness under complex backgrounds.

Rather than positioning the proposed approach as a general-purpose IQA advancement, this work emphasizes its application to AOI optical scheme evaluation and defect visibility assessment. By learning to predict engineer-consistent defect visibility scores, the framework provides a data-driven tool to assist optical configuration evaluation and reduce reliance on purely empirical manual debugging. Such quantitative modeling can provide quantitative guidance for future optimization of imaging conditions and support intelligent manufacturing processes, including coordinated control with robotic light-source adjustment systems.

The major contributions of this paper are summarized as follows:

- We formulate AOI optical scheme selection as a defect visibility regression problem based on engineer-annotated scores, providing a quantitative framework tailored to semiconductor inspection scenarios.
- We construct the WAFER-OS100 dataset, containing 9100 wafer images captured under 100 optical configurations with corresponding manual visibility ratings, supporting data-driven research in AOI optical evaluation.
- We develop an AOI-oriented multi-domain feature fusion framework (MDTNet) that integrates spatial features, frequency-domain representations, and optical scheme information to enhance sensitivity to subtle industrial defects.
- Experimental results demonstrate strong agreement between model predictions and engineer evaluations, validating the effectiveness of the proposed approach for practical AOI optical scheme assessment.

2. Related work

2.1. AOI-based industrial inspection

Automatic Optical Inspection (AOI) has been widely applied in modern industrial manufacturing as an effective vision-based inspection technique. Compared with manual inspection, AOI systems provide higher efficiency, better repeatability, and reduced human subjectivity.

AOI technologies have been extensively used in different industrial domains. In electronic manufacturing, AOI systems are commonly employed for defect detection on printed circuit boards and electronic assemblies [19,24–26]. In semiconductor manufacturing, AOI plays a crucial role in wafer surface inspection and micro-defect detection, where extremely small defects may significantly affect device yield [27,28]. In addition, AOI techniques have also been applied to mechanical surface inspection and industrial component quality assessment [29–32].

Existing AOI studies mainly focus on improving defect detection algorithms and inspection accuracy. However, relatively fewer works investigate the optimization of optical imaging conditions. In practical industrial systems, optical schemes are often determined empirically through repeated experiments conducted by experienced engineers. Such trial-and-error procedures can be time-consuming and lack quantitative evaluation mechanisms, which limits the efficiency of AOI system deployment and optimization.

2.2. Image quality assessment

NR-IQA has developed rapidly. Before the rise of deep learning, Natural Scene Statistics (NSS) theory dominated the field of IQA [33,34]. This theory assumes that natural images follow a certain statistical distribution, and various distortions disrupt this statistical regularity. Based on this theory, various handcrafted features were proposed in different domains, including spatial and gradient features [12]. However, this approach fails to model real-world distortions effectively. As a result, many studies shifted towards using deep learning for NR-IQA. With the improvement of more advanced network architectures, NR-IQA methods have evolved, from deep belief networks [35] to CNNs [28], followed by deeper CNNs [36,37], ResNet [38], and now Transformer-based models [12,14,39,40]. In addition to these developments, there are other notable works in NR-IQA. For instance, the NIMA model, designed by Talebi and Milanfar [41], replaces simple score regression with the prediction of image quality rating distributions. Ke et al. [42] proposed hash-based two-dimensional spatial embedding and scale embedding methods, enabling the model to handle input images with different resolutions and aspect ratios, while effectively integrating multi-scale information. Golestaneh et al. [12] used CNNs and Transformers to extract

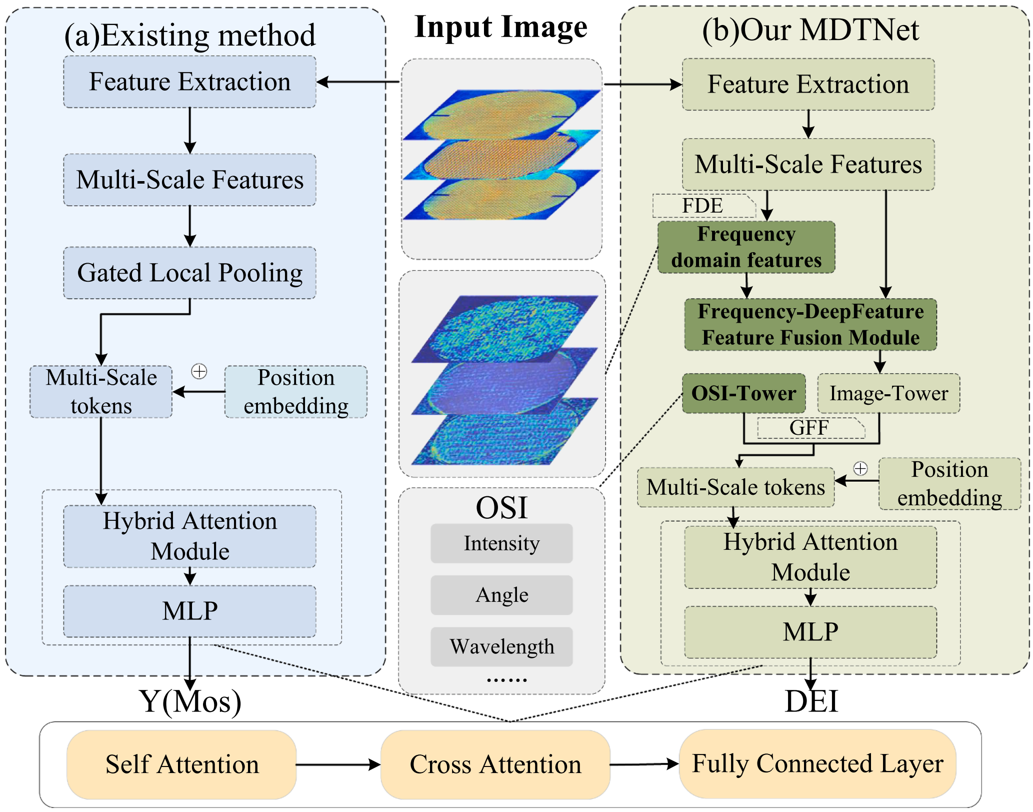


Fig. 1. Illustration of the image quality assessment flow: (a) The existing transformer-based IQA method (TOPIQ). (b) Our proposed MDTNet.

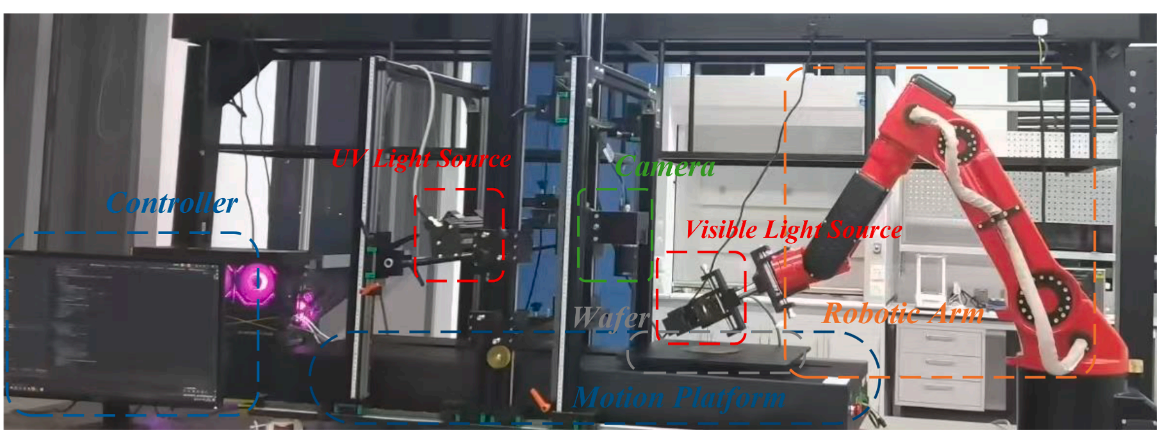


Fig. 2. Data collection equipment for defect detection: Motion platform, UV and visible light sources, robotic arm, line scan camera, and controller.

global and local features, introducing a correlated ranking loss function. These NR-IQA methods primarily use CNN or Transformer architectures to extract complex high-level features from images, evaluating overall image quality through adaptive analysis and aggregation of features.

In many practical applications, such as image communication systems and the AOI system we study, reference images are unavailable, and quality assessment relies solely on the test image itself. Therefore, NR-IQA methods are necessary for evaluating AOI image quality.

2.3. Feature extraction based on frequency domain

In the field of IQA, existing methods can generally be divided into spatial-domain feature extraction methods [34,43] and frequency-domain feature extraction methods [17]. Most existing IQA methods rely on spatial-domain features such as texture and edges, which perform well when evaluating the overall quality of an image. However, with increasing image quality requirements, particularly when handling detailed defects and complex backgrounds, spatial-domain methods have certain limitations in detecting subtle defects. Therefore, frequency-domain feature extraction methods have gradually attracted the attention of researchers. Traditional frequency-domain methods, such as Discrete Cosine Transform (DCT) [33], have been widely used in image quality assessment. These methods effectively represent distortions in the image (such as noise, blurring, etc.) by transforming the image into the frequency domain. With the development of deep learning technologies, deep learning-based frequency-domain feature extraction methods have emerged, where Fourier convolution [25,44] introduces Fourier transforms into neural networks, enabling frequency-domain operations to be directly integrated into the model training process. Fourier convolution modifies certain values in the frequency domain, thereby affecting the feature representation of the entire image, leading to a broader

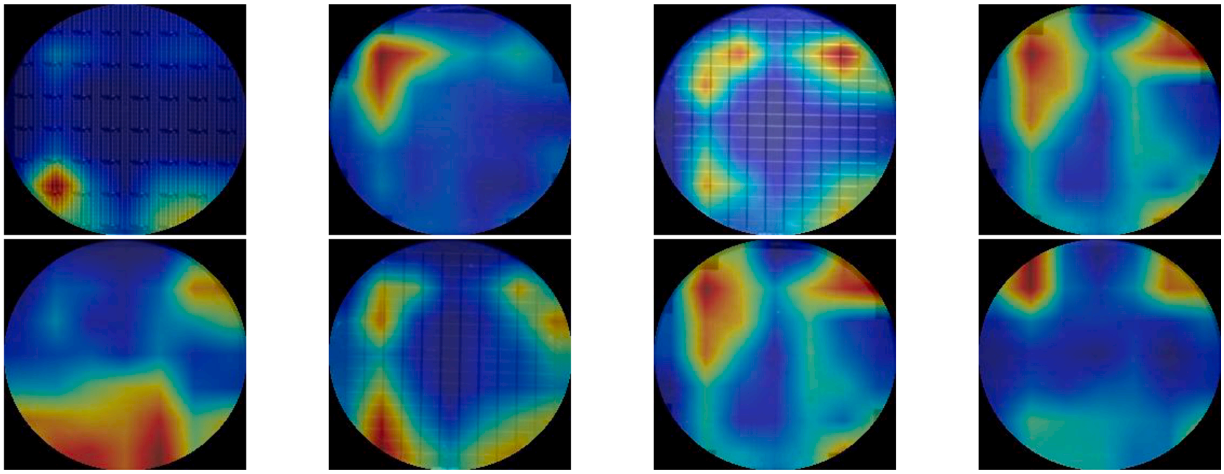


Fig. 3. Heatmaps for visualizing defects across different wafers.

receptive field and capturing global information and relationships between different components of the image. Wavelet Transform [45,46], as a frequency domain analysis method, can effectively decompose images at multiple scales, thereby capturing the detailed features of the image more finely. In AOI systems, it has significant advantages in recognizing subtle defects in complex backgrounds.

Currently, few methods in IQA combine both spatial-domain and frequency-domain features. With the introduction of deep learning, advanced methods such as Fourier convolution and Wavelet Transform can more effectively capture multi-scale and multi-directional detailed features in images, significantly improving the accuracy and robustness of image quality assessment.

3. Dataset

3.1. Dataset overview

Wafer defects are often caused by small scratches and chemical residues during processing. Due to material properties, surface optical effects, subtle defects, and complex textures, designing AOI optical schemes historically required considerable manual effort.

To study and quantify defect visibility under different optical schemes, we collected the WAFER-OS100 dataset. The dataset contains images of 91 wafers captured under 100 optical schemes, with variations in light intensity and illumination angles, producing a total of 9100 high-resolution images (2048×2048 pixels). Each image is annotated by experienced engineers with a numerical score reflecting defect visibility. Higher scores indicate better exposure of defect areas. These annotations provide a quantitative reference for comparing optical schemes and evaluating image quality under industrial inspection conditions. The resolution of WAFER-OS100 is significantly higher than mainstream IQA datasets such as LIVE, CLIVE, CSIQ, PieAPP, FLIVE, PIPAL, and KonIQ-10k, as summarized in Table 1.

Table 1
Compared with other IQA-related datasets, WAFER-OS100 has a higher resolution.

Dataset	Image numbers	Original size ($W \times H$)
LIVE [20]	779	768×512
CLIVE [47]	1162	500×500
CSIQ [21]	866	512×512
PieAPP [48]	20K	512×384
FLIVE [49]	40K	500×500
PIPAL [50]	29K	512×512
KonIQ-10k [22]	10.1K	512×384
WAFER-OS100	9.1K	2048×2048

3.2. Data acquisition equipment

The dataset was captured using a high-precision programmable optical acquisition system in Fig. 2, which includes a linear motion platform, an industrial line-scan camera, a tunable light source system, and a six-degree-of-freedom robotic arm. The motion platform coordinates with the camera to acquire wafer images, while the robotic arm adjusts the illumination angle for multi-angle acquisition. The light source system allows programmable intensity control and includes both visible and UV light sources.

During data acquisition, a UV fluorescence labeling method was employed to identify typical defects such as scratches, edge damage, and surface pits. Fluorescent marking was only used during data acquisition and manual annotation to help engineers identify defect regions more accurately. The fluorescence information was not used as an input to the network or as an auxiliary supervision signal during training, it does not affect the scoring. Images were captured under multi-angle lighting by moving the robotic arm in 5° steps in the range of 22.5° to 67.5°, with varying light intensities.

3.3. Optical scheme evaluation

Each image in WAFER-OS100 was independently evaluated by three experienced AOI engineers under visible-light inspection conditions. In Table 2, the annotators assessed the visibility of defect regions using a five-level DEI scoring criterion, where higher scores indicate clearer defect exposure and better visibility under a given optical scheme. Representative intermediate feature responses under different optical schemes are illustrated in Fig. 3. The heatmaps are generated from intermediate activation responses of the backbone feature extractor in MDTNet and are used for qualitative visualization of defect-sensitive regions. To ensure annotation consistency, all annotators followed unified evaluation guidelines considering defect clarity, boundary distinguishability, and local contrast around defect regions.

The final DEI label for each image was obtained by averaging the scores from all annotators. For samples exhibiting large inter-rater disagreement, a joint review discussion was conducted to determine the final consensus score. In addition, inter-rater reliability was evaluated using the Intraclass Correlation Coefficient (ICC). The computed ICC value was 0.87, indicating good annotation consistency.

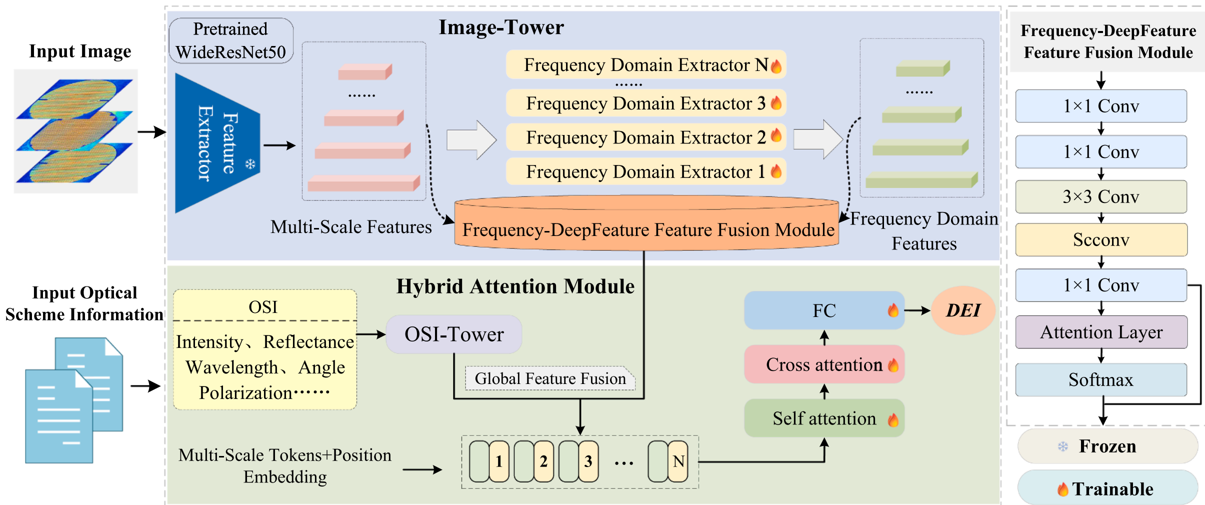


Fig. 4. Architecture overview of the proposed MDTNet.

Table 2
Defect visibility scoring criterion used in WAFER-OS100.

Score	Meaning
5	Defect region is clearly distinguishable with sharp boundaries
4	Defect region is clearly observable
3	Defect region is observable with reduced boundary clarity
2	Defect region is weakly observable
1	Defect region is barely identifiable

4. Method

To solve the problem that small defects or background noise are often ignored in AOI applications, this chapter introduces in detail the innovative method of combining the Wavelet Transform and Transformer. The overall architecture is shown in Fig. 4, WideResNet50 [23] is first used to extract multi-scale features of the image. The FDE module can better capture the detailed information of the image by enhancing the expression ability of frequency domain features, and uses Transformer to model the long-range dependencies between different frequency components. Through the FDFFM module, multi-domain feature fusion is performed. We can not only identify small defects more effectively, but also increase the focus on global information and improve the accuracy of image quality assessment. This method provides a new solution for dealing with complex backgrounds and small defects, overcoming the shortcomings of existing methods. At the same time, the image OSI domain is fused to further improve the accuracy of the AOI system in practical applications.

4.1. Frequency domain extractor(FDE)

Our method proposes a wavelet-domain complex convolution feature enhancement module FDE. The specific structure of the Module is shown in Fig. 5. Given an input feature map $X \in \mathbb{R}^{B\times C\times H\times W}$, we first decompose it into a low-frequency component $X_l \in \mathbb{R}^{B\times C\times H/2\times W/2}$ and high-frequency complex subbands $X_h \in \mathbb{C}^{B\times 3\times C\times H/2\times W/2}$ using a 2D wavelet transform Wav, defined as:

$$X_l, X_h = Wav(X) \tag{1}$$

where X_l is the low-frequency component, and X_h represents the high-frequency complex subbands capturing detailed information in horizontal, vertical, and diagonal directions.In Eq. (1), Wav denotes the Dual-Tree Complex Wavelet Transform (DT-CWT) with J = 2 decomposition levels. We further decompose X_h into its real part X_r and imaginary part X_i:

$$X_r = Re(X_h), \quad X_i = Im(X_h) \tag{2}$$

where Re(·) and Im(·) denote the real and imaginary parts of the complex subbands, respectively. After reshaping and block-wise partitioning X_r and X_i, we perform feature interaction through a two-stage complex convolution, Stage1 has formulas such as Eqs. (3) and (4), and Stage2 has formulas such as Eqs. (5) and (6):

$$X_{r_1} = ReLU(W_{r_1}X_r - W_{i_1}X_i + b_{r_1}) \tag{3}$$

$$X_{i_1} = ReLU(W_{i_1}X_r + W_{r_1}X_i + b_{i_1}) \tag{4}$$

$$X_{r_2} = W_{r_2}X_{r_1} - W_{i_2}X_{i_1} + b_{r_2} \tag{5}$$

$$X_{i_2} = W_{i_2}X_{r_1} + W_{r_2}X_{i_1} + b_{i_2} \tag{6}$$

where W are the weight matrices for linear transformations;b are the bias vectors;X_1 are the outputs of the first layer after applying the ReLU activation function; and X_2 are the final outputs of the second layer. We then apply adaptive threshold shrinkage to the output:

$$X_h^{out} = Softshrink(X_{r_2} + jX_{i_2}, \lambda) \tag{7}$$

where the threshold λ is dynamically generated through global average pooling of the feature magnitudes. Finally, the processed low-frequency component X_l and high-frequency component X_h^{out} are fused and reconstructed through an inverse wavelet transform W^{-1} to obtain x_{out}:

$$X_{out} = Wav^{-1}(X_l, X_h^{out}) \tag{8}$$

4.2. Frequency-deep feature feature fusion module(FDFFM)

Building on previous research on spatial shift operations (ShiftConv) [51] and attention mechanisms (Attn)[52], we propose an innovative lightweight feature processing module. This module combines ShiftConv and the Attn mechanism to enhance feature extraction and aggregation in super-resolution tasks. Specifically, ShiftConv effectively aggregates local features through spatial shift operations, while the Attn mechanism can automatically focus on key parts within multiscale features, thus improving the model's ability to perceive image details. We have further optimized the integration of the two, enabling it to more effectively extract and fuse critical features in low-level image restoration tasks while maintaining computational efficiency.

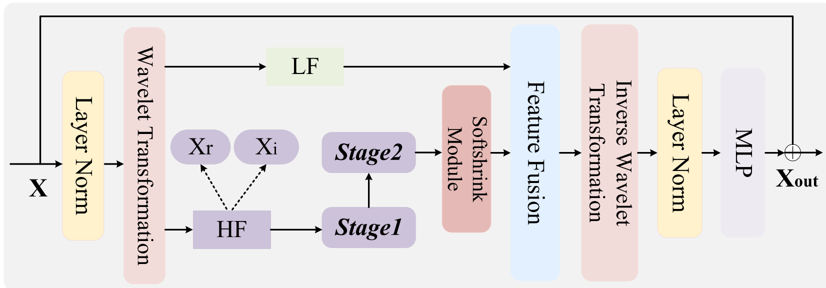


Fig. 5. Our FDE module, used to enhance the representation of high-frequency information.

Traditional convolutional neural networks typically rely on computationally intensive convolution operations, such as 3×3 convolution blocks (ResBlocks) [51], to extract and aggregate features. However, these operations have certain limitations in reducing computational costs and improving efficiency. To effectively reduce the computational overhead while maintaining good reconstruction performance, we propose a lightweight feature extraction module by combining spatial shift operations with 1×1 convolutions. In this module, the input features $f \in \mathbb{R}^{C_{\text{lat}}\times H\times W}$ are first evenly divided into n times smaller feature groups $f_i \in \mathbb{R}^{C_{\text{lat}}/n\times H\times W}$ along the channel dimension. Then, a spatial shift operation is applied to each separated feature group, resulting in the shifted features $f_{shift} \in \mathbb{R}^{C_{\text{lat}}/n\times H\times W}$. This operation explores the relationships between local pixels in the feature map and aggregates them using 1×1 convolutions, significantly reducing the computational overhead. The spatial shift operation can be represented as:

$$f_{shift} = ShiftConv(f_i, step_i) \tag{9}$$

where step_i represents the shift step, defined as $\text{step}_i = (d_h \cdot s_h, d_w \cdot s_w)$, where d_h and d_w are the shift direction, s_h and s_w are the step size, respectively. With these step parameters, we can perform feature aggregation at different spatial locations.

To further enhance the model's ability to focus on key features, we introduce an attention mechanism (Attn) to guide the feature aggregation process. This mechanism calculates the similarity between the query and key vectors and weights the output value vector based on the similarity, guiding the model to focus on key features within the image. Specifically, the feature vectors for the query, key, and value are generated by learnable weight matrices, and the attention output is calculated based on these features, thereby optimizing the image quality assessment. The mechanism calculates the similarity between the query and key through a dot product, generates a weighted matrix, and applies it to the value vector to produce the weighted output. In this process, we generate queries (Q), keys (K), and values (V) for each spatially shifted feature f_{shift}:

$$Q = W_q f_{shift}, \quad K = W_k f_{shift}, \quad V = W_v f_{shift} \tag{10}$$

where W_q , W_k , W_v is a logical weight matrix. Then, the attention weights are generated by calculating the similarity between the queries and keys, and the value features are weighted based on these weights:

$$Attn(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

where d_k is the key vector dimension. This attention mechanism helps the network adaptively focus on the regions of the image that are most important for the super-resolution task, thereby enhancing the feature representation capability.

4.3. OSI fusion module

In this study, we investigated a method that combines Optical Scheme Information (OSI) with image features to enhance model performance. And its specific structure is shown in the Fig. 6. Unlike conventional NR-IQA methods that rely solely on image input, the proposed framework additionally incorporates physically meaningful optical parameters, including illumination intensity and incident angle, which are naturally available in practical AOI imaging systems rather than artificially introduced auxiliary labels. The OSI branch does not use wafer identifiers or optical scheme IDs. Instead, only physically meaningful optical parameters are used as inputs. Therefore, the model is encouraged to learn the relationship between imaging conditions and defect visibility rather than relying on discrete optical scheme identifiers. These optical parameters are further encoded together with image features to facilitate adaptive defect visibility estimation under different imaging conditions. To ensure fair comparison, we additionally evaluate a variant without OSI input in the ablation study. The overall process is outlined as follows:

First, the OSI is processed via a Multilayer Perceptron (MLP) [53] to generate context embeddings. Let $\text{OSI} \in \mathbb{R}^{d_{\text{OSI}}}$ denote the input OSI. The MLP consists of two hidden layers. The outputs of these layers are given by:

$$h_1 = ReLu(W_1 OSI + b_1) \tag{12}$$
$$Attn_{OSI} = Sigmoid(W_2 h_1 + b_2) \tag{13}$$

where W_1, W_2 are weight matrices, b_1, b_2 are bias vectors and $Attn_{OSI}$ is the attention weight of OSI domain.

Next, these embeddings are combined with the Fused Frequency-DeepFeature Feature $\text{F} \in \mathbb{R}^{d_{\text{feature}}}$ through linear transformations and global multiplication:

$$h_3 = W_f F + b_f \tag{14}$$
$$F_{fusion} = h_3 \odot Attn_{OSI} \tag{15}$$

where W_f is weight matrice, b_f is bias vector, h_3 is the output of the hidden layer and F_{fusion} is the fused feature.

4.4. Hybrid attention module

In terms of feature enhancement, the multiscale feature representation obtained by extraction of features in the frequency domain is denoted as $F_{fusion_1}, \ldots, F_{fusion_i} \in \mathbb{R}^{(D+E)\times H_i\times W_i}$, where $D+E$ is the number of feature channels, and H_i and W_i are the height and width of the feature map respectively, $i \in \{1,2,\ldots,n\}$ and $\text{n} = 5$ for WideResNet50. These features capture both global and local information of the image.

To further enhance the representation ability of these features, we apply a self-attention (SA) mechanism to enhance F_{fusion_i}.

Before this, we first incorporate position encoding to preserve the spatial information of the features. The position encoding $P \in \mathbb{R}^{(D+E)\times H_i\times W_i}$ can be generated using sine and cosine functions or implemented as learnable parameters. The feature representation after adding position encoding is denoted as:

$$F^{pos}_{fusion_i} = F_{fusion_i} + P \tag{16}$$

Subsequently, the SA is computed based on the following formulation:

$$Q = W_Q F^{pos}_{fusion_i}, \quad K = W_K F^{pos}_{fusion_i}, \quad V = W_V F^{pos}_{fusion_i} \tag{17}$$

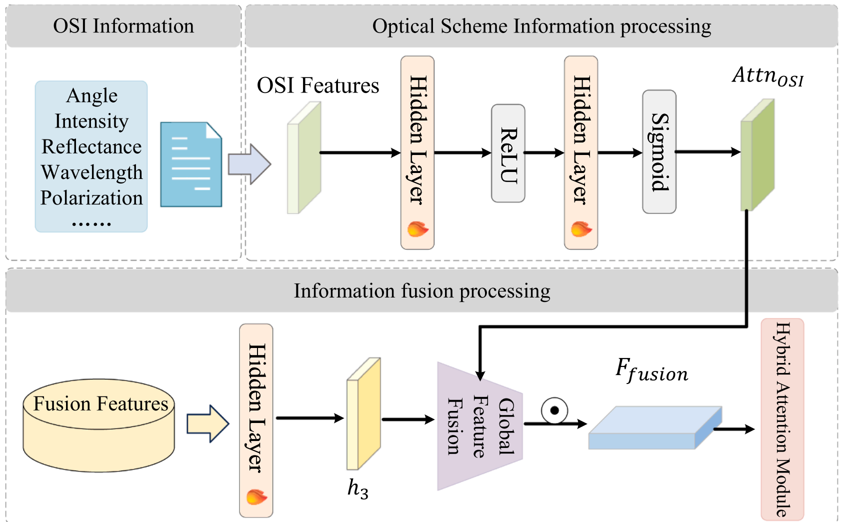


Fig. 6. Our OSI module: Fusion of optical scheme information.

$$F'_{fusion_i} = SA(F_{fusion_i}) = Attn(Q, K, V) + F_{fusion_i} \quad (18)$$

where W_Q ,W_K and W_V are learnable weight matrices, and F'_{fusion_i} is the enhanced feature.

The self-attention mechanism can capture the relationships between multi-scale features, thereby enhancing the model's understanding of global context and further improving the accuracy of image quality assessment.

To further exploit the relationships between multi-scale features, we introduce a Cross-Scale Attention (CSA) mechanism on top of the self-attention mechanism. The cross-scale attention mechanism computes attention scores between features at different scales, enabling effective fusion of multi-scale information. Given the enhanced features $F'_{fusion_1}, \dots, F'_{fusion_i} \in \mathbb{R}^{(D+E)\times H_i \times W_i}$ from multiple scales, the cross-scale attention operation is defined as:

$$F''_{fusion_j} = CSA(F'_{fusion_1}, \dots, F'_{fusion_i}) \quad (19)$$

where CSA denotes the cross-scale attention function.

Finally, F''_{fusion_j} will be mapped to quality scores through a MLP, $j \in \{1, 2, \dots, n-1\}$, as shown in the following formula:

$$\hat{y} = MLP(F''_{fusion_j}) \quad (20)$$

5. Experiments

In this section, the conducted experiments are described and evaluated in detail. First, the datasets, evaluation metrics, and implementation details required for the experiments are introduced. Then, qualitative and quantitative results of the MDTNet approach and the current state-of-the-art methods on the wafer dataset constructed in this paper and several public datasets are reported. Finally, through ablation experiments, the impact of each component of the MDTNet is analyzed in detail, and its scalability is evaluated based on the results.

5.1. Dataset

This paper conducts experiments on several NR-IQA public datasets, such as CLIVE, KonIQ-10k, etc., and the wafer image dataset WAFER-OS100. WAFER-OS100 contains 9100 images collected by the AOI system under 100 different optical schemes.

To avoid potential data leakage, the dataset was split at the wafer level rather than the image level. Specifically, all images captured from the same wafer under different optical schemes were strictly assigned to the same subset. Therefore, no wafer appeared simultaneously in both the training and testing sets. This protocol ensures that the model generalizes across unseen wafers instead of memorizing wafer-specific appearance patterns. The detailed dataset split statistics are summarized in Table 3.

Table 3
Dataset split statistics of WAFER-OS100.

Split	Wafers	Images
Train	73	7300
Test	18	1800

5.2. Evaluation metrics

Two common metrics are used to evaluate the performance of the IQA method proposed in this paper: Spearman Rank Correlation Coefficient (SRCC) and Pearson Linear Correlation Coefficient (PLCC), which have been widely used in previous IQA methods. SRCC evaluates the accuracy of the predicted ranking of image quality, while PLCC measures the correlation between the predicted score and the true score. In this paper, the correlation between the true score of the predicted wafer and the score predicted by the NR-IQA model is calculated. The formula of $PLCC$ is as follows:

$$PLCC = \frac{\sum_{i=1}^{N}(y_i - u_y)(\hat{y}_i - u_{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - u_y)^2} \cdot \sqrt{\sum_{i=1}^{N}(\hat{y}_i - u_{\hat{y}})^2}} \quad (21)$$

where y_i and $\hat{y}_i$ are the ground-truth quality score and predicted quality score of the ith image, respectively, and u_y and $u_{\hat{y}}$ are the mean values of all ground-truth scores and predicted scores, respectively.The formula of $SRCC$ is as follows:

$$SRCC = \frac{\sum_{i=1}^{N}(r_{y_i} - \bar{r}_y)(r_{s_i} - \bar{r}_s)}{\sqrt{\sum_{i=1}^{N}(r_{y_i} - \bar{r}_y)^2} \cdot \sqrt{\sum_{i=1}^{N}(r_{s_i} - \bar{r}_s)^2}} \quad (22)$$

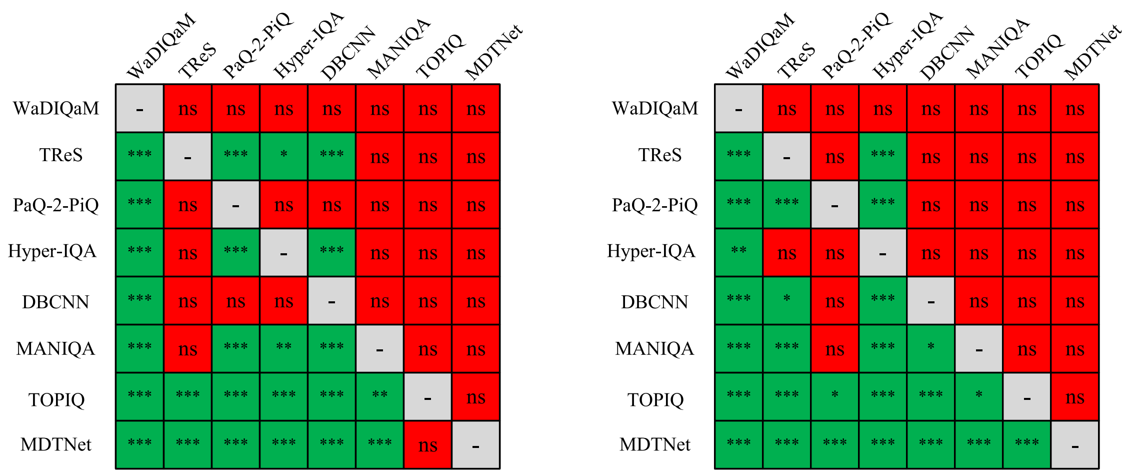
where r_{y_i} and r_{s_i} represent the rank positions of the ith image's ground-truth quality score and predicted quality score, respectively, while $\bar{r}_y$ and $\bar{r}_s$ denote the mean ranks across all ground-truth scores and predicted scores, respectively.

5.3. Implementation details

All experiments were conducted in the same hardware environment, using NVIDIA 4080Ti GPU for training and inference. For each dataset,

Table 4
Performance comparison on public benchmarks and our dataset.

Method	CLIVE		KonIQ-10K		WAFER-OS100		FLOPs (G)
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	
WaDIQaM [54]	0.682	0.671	0.804	0.807	0.7878	0.7824	11.13
TReS [12]	0.864	0.877	0.915	0.928	0.8106	0.8112	419.36
PaQ-2-PiQ [49]	0.84	0.85	0.87	0.88	0.8305	0.8216	16.19
Hyper-IQA [55]	0.859	0.882	0.906	0.917	0.7777	0.7635	108.39
DBCNN [37]	0.869	0.869	0.875	0.884	0.8220	0.8105	64.65
MANIQA [14]	0.883	0.909	0.928	0.939	0.8164	0.8249	572.34
TOPIQ [16]	0.870	0.884	0.926	0.939	0.8464	0.8380	27.27
Ours	0.866	0.881	0.930	0.943	0.8732	0.8840	73.28
Ours(+ OSI)	–	–	–	–	**0.9069**	**0.8955**	75.72



(a) KonIQ-10k (b) WAFER-OS100

Fig. 7. Statistical significance tests on KonIQ-10K and WAFER-OS100 datasets. The '***' represents that the method in that row is extremely significantly superior to the method in the corresponding column. The '**' means that the method in that row is very significantly superior to the corresponding column method. The symbol '*' means that the method in this row is significantly superior to the corresponding column method. The 'ns' means that the method in that row is not significantly superior to the method in the corresponding column. The symbol '-' means that the two methods in the corresponding row and column are statistically indistinguishable.

the network was trained based on the model, using the AdamW optimizer, with an initial learning rate of 3e-5 and a step decay strategy. The model training process lasted for 100 epochs with a batch size of 16. To ensure the generalization ability of the model, we used data augmentation techniques during training, including image rotation, flipping, and random cropping.

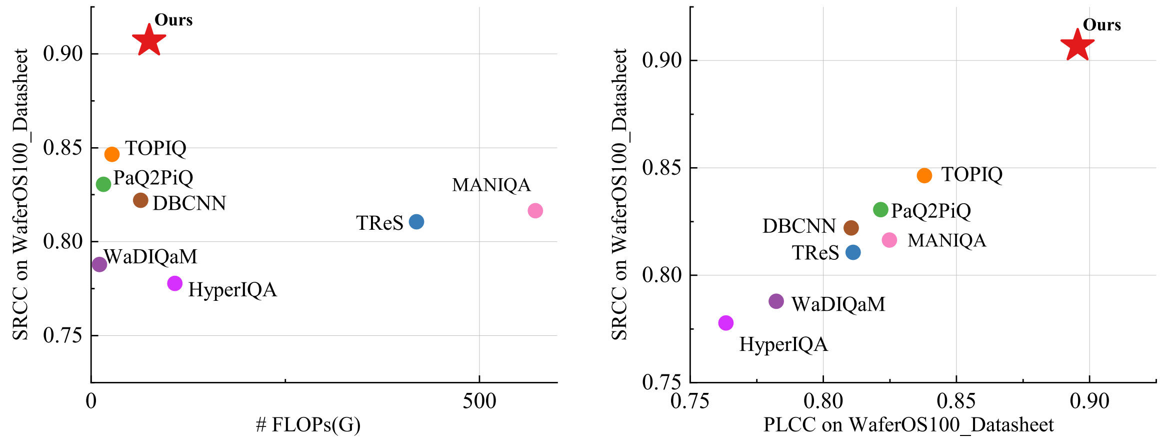
5.4. Comparisons with state-of-the-arts

For other state-of-the-art methods in NR-IQA in recent years such as WaDIQaM [54], TReS [12], PaQ-2-PiQ [49], Hyper-IQA [55], DBCNN [37], and TOPIQ [16], our proposed method is comprehensively compared. The results are shown in the following table: The experimental results show that the Multi-Domain Transformer model proposed in this paper outperforms the existing methods in PLCC and SRCC evaluation indicators. The results are as in Table 4. At the same time, we conducted statistical significance tests and used the left-tailed F-test to perform a series of hypothesis tests on the residuals between the subjective and objective quality scores obtained from each NR-IQA method after nonlinear regression, as well as the residuals between the true score obtained from the MDTNet method and the predicted score. The statistically significant results of the KonIQ-10K and WAFER-OS100 datasets are shown in Fig. 7a and 7b, respectively. From the results, it can be seen that our MDTNet is not inferior to other IQA methods on the publicly available dataset KonIQ-10K, and performs the best on the WAFER-OS100 dataset, significantly better than other IQA methods. In particular, on the WAFER-OS100 dataset, our method combines frequency domain features with multi-scale features and integrates OSI domain features for the first time, showing excellent performance in detail defect detection, as shown in the following Fig. 8. Compared with existing methods, SRCC and PLCC improvements of 6.05% and 5.75% were achieved at a small number of FLOPS, which can more accurately identify subtle defects in wafer images. In addition, for public datasets that focus on the overall quality of images, our method also performs well and can effectively process high-frequency details and evaluate overall images.

5.5. Ablation studies

In this section, we first analyze the effects of different feature extraction backbones on NR tasks, and then present the ablation experiments on the proposed components in MDTNet.

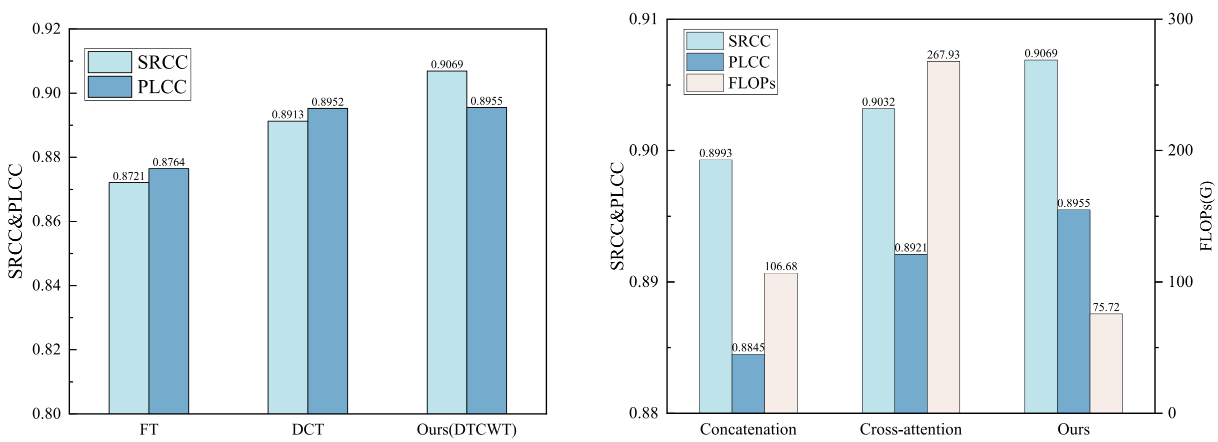
Performances with different backbones. Different backbones have a significant impact on the performance of image quality estimation. Therefore, we further evaluate how different backbones affect the performance of NR on the WAFER-OS100 dataset. In the experiment, we first compare ResNet50 [56] and WideResNet50 [23], two kinds of pretrained backbones used for feature extraction, and the results are



(a) SRCC vs. FLOPs

(b) SRCC vs. PLCC

Fig. 8. Comparison among SOTA IQA methods on WAFER-OS100 dataset. Fig. (a) shows the relationship between computational cost (Flops) and performance (SRCC) of the MDTNet NR task. Fig. (b) shows the relationship between SRCC and PLCC.



(a) Performance vs. Frequency-Band Selection

(b) Performance and FLOPs vs. OSI-Fusion Strategy

Fig. 9. The results of different Ablation Study: (a) The impact of different frequency domain feature extraction methods on the performance of the MDTNet model; (b) The impact of different information fusion strategies for Optical Scheme Information on the performance of MDTNet models.

shown in Table 5. We can observe that on the WAFER-OS100 dataset, the performance improvement brought by using WideResNet50 as the backbone is more obvious than using ResNet50, with SRCC increased by 0.0221 and PLCC increased by 0.0409. We assume that the main reason is that WideResNet50 can extract more and more kinds of feature information in each layer by increasing the number of channels in each residual block, to capture the differences of more scales and details, and improve the expression ability of multi-scale information.

Ablation of the proposed components. In Table 5, we evaluate the components proposed in MDTNet on the WAFER-OS100 dataset. It should be noted that the self-attention and cross-scale attention mechanisms are inherited from the original TOPIQ framework rather than newly proposed components in this work. Therefore, our ablation study mainly focuses on evaluating the contributions of the newly introduced modules, including FDE, FDFFM, and OSI. The baseline models are ResNet50 and WideResNet50 pre-trained networks with multi-scale features. In particular, the baseline configuration in Table 5 with Backbone ① corresponds to the original TOPIQ framework without the proposed FDE, FDFFM, and OSI modules. Therefore, the resulting SRCC and PLCC values are identical to those of TOPIQ in Table 4. We evaluate three components of MDTNet: 1) FDE; 2) FDFFM; and 3) OSI. Experimental results show that all three components contribute positively to the final performance. Specifically, after introducing the FDE module, both PLCC and SRCC are improved, indicating that frequency-domain representations are beneficial for capturing high-frequency defect details and subtle structural variations. The OSI module brings the most significant performance improvement, demonstrating that defect visibility assessment is influenced not only by image content itself but also by optical imaging conditions such as illumination intensity and incident angle. This observation highlights the importance of incorporating optical scheme information into AOI-oriented NR-IQA tasks. FDFFM is designed as a feature interaction and fusion module operating on the frequency-domain representations generated by FDE, rather than as a completely standalone component. Therefore, its contribution is evaluated incrementally on top of the FDE-enhanced baseline.Although the performance gain brought by FDFFM is relatively moderate compared with OSI, it consistently improves the interaction between frequency-domain representations and deep spatial features, contributing to more stable defect visibility estimation. The complete MDTNet framework achieves the best overall performance, indicating that the combination of frequency-domain features, feature fusion mechanisms, and optical scheme information effectively improves both global representation capability and fine-grained defect perception.

Ablation study on frequency band selection. To validate the choice of frequency domain representation, we compare the performance of Dual Tree Complex Wavelet Transform (DTCWT) with Fourier Transform (FT) and Discrete Cosine Transform (DCT) for feature extraction. As shown in Fig. 9a, DTCWT achieves superior

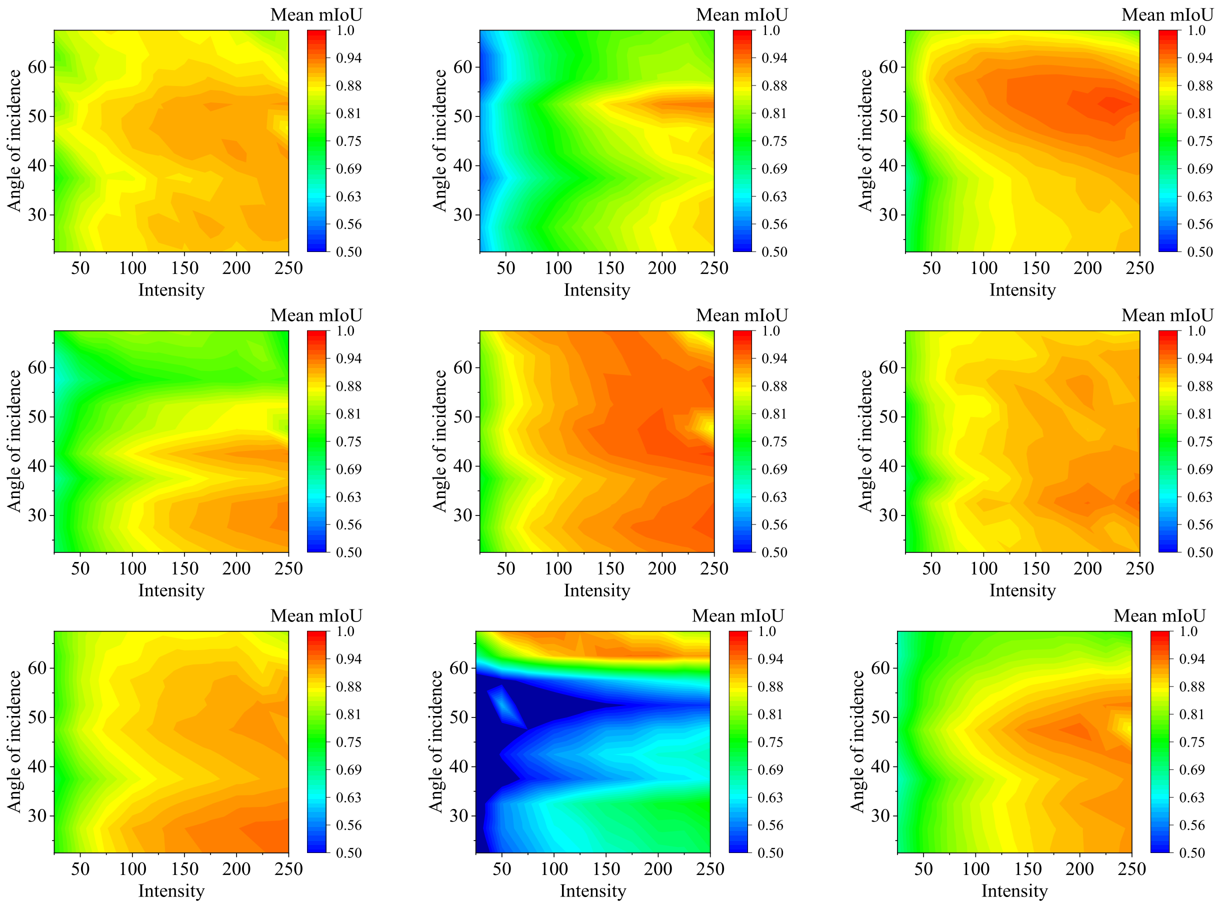


Fig. 10. Downstream defect segmentation performance under different optical schemes.

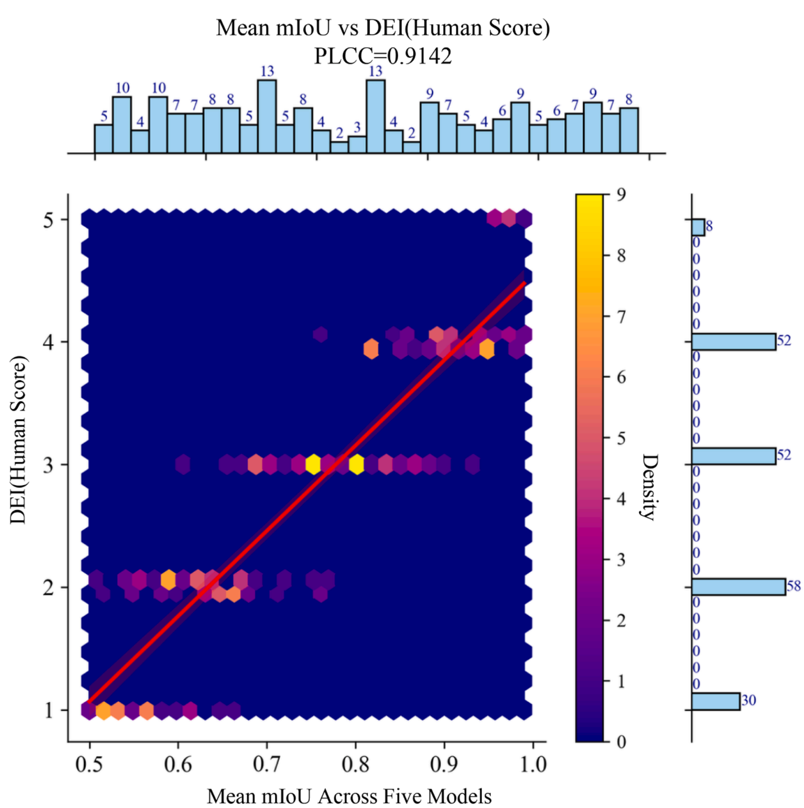


Fig. 11. Correlation between averaged mIoU and manual DEI scores under different optical configurations.

Table 5
Ablation study through WAFER-OS100 dataset experiments for different components in MDTNet. Backbone ① corresponds to the original TOPIQ framework with a ResNet50 backbone and ② is WideResNet50.

FDE	FDFFM	OSI	Backbone	SRCC	PLCC
–	–	–	①	0.8464	0.8380
✓	–	–	①	0.8570	0.8549
–	–	✓	①	0.8654	0.8594
✓	✓	–	①	0.8595	0.8506
✓	✓	✓	①	**0.8957**	**0.8851**
–	–	–	②	0.8685	0.8789
✓	–	–	②	0.8815	0.8860
–	–	✓	②	0.8935	0.8914
✓	✓	–	②	0.8832	0.8840
✓	✓	✓	②	**0.9069**	**0.8955**

performance on the WAFER-OS100 dataset compared to FT and DCT. This improvement can be attributed to the ability of DTCWT to decompose the image into local frequency components, which enhances the detection of multi-scale defects and preserves spatial frequency details more effectively than global transforms such as FT and DCT. These findings justify our choice of wavelet-based decomposition for defect-sensitive feature extraction in AOI systems.

Ablation study on OSI fusion strategy. To evaluate the effectiveness of our OSI fusion scheme, we compare it with other methods: (1) direct cascading and (2) criss-cross attention. As shown in Fig. 9b, our method achieves a better balance between efficiency and accuracy, improving SRCC by 0.37% and PLCC by 0.34% compared to cross-attention, while reducing the computational cost. Although cross-attention provides powerful feature interactions, its high complexity limits its real-time application in industrial settings. In contrast, our lightweight multi-layer perceptron-based fusion method is more computationally efficient while maintaining competitive advantages, making it more suitable for practical AOI deployments.

Ablation study on softshrink threshold optimization. To verify the effectiveness of the adaptive soft shrinkage threshold λ in Eq. (7), we compared it with different fixed thresholds λ in Table 6. Experiments on the WAFER-OS100 dataset show that our adaptive λ achieves better results than the fixed threshold while being more robust in suppressing high-frequency noise. The adaptive mechanism dynamically adjusts λ according to the feature amplitude, filtering out noise while retaining key defect features. Ablation confirms that our method achieves the best balance between noise suppression and feature preservation for AOI tasks.

5.6. Downstream defect segmentation validation

To further validate the practical relevance of the proposed DEI evaluation framework for AOI applications, we additionally conducted downstream defect segmentation experiments under different optical schemes.

Specifically, five semantic segmentation models, including DeepLabV3, DeepLabV3+, FCN(HRNet), PSPNet, and BiSeNetV1, were evaluated under different illumination conditions, and the averaged mIoU values were calculated for each optical configuration. As shown in Fig. 10, different optical schemes exhibit significantly different defect exposure characteristics, leading to noticeable variations in downstream defect segmentation performance.

Furthermore, we analyzed the correlation between the manually annotated DEI scores and the averaged mIoU values across different optical configurations. As illustrated in Fig. 11, the proposed DEI scores show a strong positive correlation with downstream segmentation performance.

These results indicate that optical schemes with higher DEI scores generally provide clearer defect visibility and better defect boundary representation, thereby improving downstream defect inspection capability. This observation further demonstrates that the proposed framework provides meaningful evaluation guidance for practical AOI optical assessment tasks. These results further support that DEI is not merely a subjective perceptual score, but also reflects practical downstream inspection effectiveness under different optical configurations.

Table 6
Performance comparison with fixed vs. dynamic λ.

Performance	The value of λ						
	Fixed λ						Dynamic λ
	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$	$\lambda = 10$	
SRCC	0.8877	0.8943	0.8978	0.8871	0.8914	0.9008	**0.9069**
PLCC	0.8795	0.8799	0.8810	0.8839	0.8841	0.8945	**0.8955**

6. Conclusion

In this study, we addressed the problem of optical scheme evaluation in Automatic Optical Inspection (AOI) systems from a quantitative perspective. Instead of relying solely on empirical visual comparison, we formulated defect visibility assessment as a supervised regression task based on engineer-annotated scores. This formulation provides a structured approach to modeling the relationship between optical configuration and perceived defect clarity, contributing to more systematic optical scheme assessment in semiconductor inspection. The construction of the WAFER-OS100 dataset establishes a dedicated benchmark for studying the effects of optical configurations in AOI scenarios. By incorporating multi-angle illumination settings and expert visibility ratings, the dataset supports data-driven analysis of imaging condition evaluation, which has traditionally depended on manual experience. The proposed MDTNet framework demonstrates that integrating frequency-domain representations and optical scheme information can effectively enhance sensitivity to subtle defects under complex industrial backgrounds. The strong consistency between model predictions and engineer assessments indicates that defect visibility can be learned in a stable and reliable manner, providing practical support for AOI optical evaluation. It should be noted that the present study focuses on semiconductor wafer inspection, and performance in other industrial contexts may depend on defect characteristics and surface properties. Future work will extend the dataset to additional AOI applications and further investigate reinforcement learning-based optical parameter optimization and cross-scenario adaptability to improve robustness in diverse manufacturing environments.

CRediT authorship contribution statement

Zikang Zeng: Writing – original draft, Software; **Yiru Wang:** Writing – original draft; **Jinghan Wang:** Validation; **Tianwu Lei:** Software; **Mingwei Feng:** Writing – review & editing; **Ningmu Zou:** Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.S. Saif, K. Aras, A. Giuseppi, Automated optical inspection for printed circuit board assembly manufacturing with transfer learning and synthetic data generation, in: 2022 30th Mediterranean Conference on Control and Automation (MED), IEEE, 2022, pp. 318–323.

[2] C.-F.J. Kuo, T.-y. Fang, C.-L. Lee, H.-C. Wu, Automated optical inspection system for surface mount device light emitting diodes, J. Intell. Manuf. 30 (2019) 641–655.
[3] T.-H. Sun, F.-C. Tien, F.-C. Tien, R.-J. Kuo, Automated thermal fuse inspection using machine vision and artificial neural networks, J. Intell. Manuf. 27 (2016) 639–651.
[4] S.-L. Lu, X.-M. Zhang, Y.-C. Kuang, Optimized design of an AOI illuminator, in: 2007 International Conference on Wavelet Analysis and Pattern Recognition, 2, IEEE, 2007, pp. 924–928.
[5] G. Xiong, S. Ma, X. Nie, X. Tang, New multicolor illumination system for automatic optical inspection, in: 5th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optical Test and Measurement Technology and Equipment, 7656, SPIE, 2010, pp. 1038–1046.
[6] F.-M. Tzu, J.-H. Chou, Effectiveness of light source on detecting thin film transistor, Crystals 8 (10) (2018) 394.
[7] Y. Du, J. Chen, H. Zhou, X. Yang, Z. Wang, J. Zhang, Y. Shi, X. Chen, X. Zheng, An automated optical inspection (AOI) platform for three-dimensional (3D) defects detection on glass micro-optical components (GMOC), Opt. Commun. 545 (2023) 129736.
[8] W.U. Yilong, W. Junling, P. ZHANG, Application of AOI light source modes in multichip modules inspection, in: 2018 19th International Conference on Electronic Packaging Technology (ICEPT), IEEE, 2018, pp. 141–143.
[9] H.T. Kim, S.T. Kim, Y.J. Cho, Quick and efficient light control for conventional automatic optical inspection (AOI) systems, Int. J. Precis. Eng. Manuf. 16 (2) (2015) 247–254.
[10] C.-K. Cheng, H.-Y. Tsai, Enhanced detection of diverse defects by developing lighting strategies using multiple light sources based on reinforcement learning, J. Intell. Manuf. 33 (8) (2022) 2357–2369.
[11] J. Shi, P. Gao, J. Qin, Transformer-based no-reference image quality assessment via supervised contrastive learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, 38, 2024, pp. 4829–4837.
[12] S.A. Golestaneh, S. Dadsetan, K.M. Kitani, No-reference image quality assessment via transformers, relative ranking, and self-consistency, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1220–1230.
[13] Z. Wang, A.C. Bovik, Modern Image Quality Assessment, Ph.D. thesis, Springer, 2006.
[14] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, Y. Yang, MANIQA: multi-dimension attention network for no-reference image quality assessment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1191–1200.
[15] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2013) 684–695.
[16] C. Chen, J. Mo, J. Hou, H. Wu, L. Liao, W. Sun, Q. Yan, W. Lin, TOPIQ: a top-down approach from semantics to distortions for image quality assessment, IEEE Trans. Image Process. 33 (2024) 2404–2418.
[17] X. Li, T. Jiang, H. Fan, S. Liu, SAM-IQA: can segment anything boost image quality assessment?, (2023). arXiv preprint arXiv:2307.04455
[18] Z. Chen, Q. Zhang, M. Zhong, T. Kou, C. Liu, H. Jing, G. Liu, J. Shen, Multidimensional computed measurement for highly accurate PCBA defect detection, Opt. Express 33 (3) (2025) 5340–5354.
[19] M. Abd Al Rahman, A. Mousavi, A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry, IEEE Access 8 (2020) 183192–183271.
[20] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.
[21] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 011006.
[22] V. Hosu, H. Lin, T. Sziranyi, D. Saupe, KonIQ-10k: an ecologically valid database for deep learning of blind image quality assessment, IEEE Trans. Image Process. 29 (2020) 4041–4056.
[23] S. Zagoruyko, N. Komodakis, Wide residual networks, (2016). arXiv preprint arXiv:1605.07146
[24] C.-M. Lo, T.-Y. Lin, Automated optical inspection based on synthetic mechanisms combining deep learning and machine learning, J. Intell. Manuf. 36 (2024) 1–15.
[25] S.-H. Chen, D.-B. Perng, Automatic optical inspection system for IC molding surface, J. Intell. Manuf. 27 (2016) 915–926.
[26] T. Parakontan, W. Sawangsri, Development of the machine vision system for automated inspection of printed circuit board assembl, in: 2019 3rd International Conference on Robotics and Automation Sciences (ICRAS), IEEE, 2019, pp. 244–248.
[27] S.-H. Huang, Y.-C. Pan, Automated visual inspection in the semiconductor industry: a survey, Comput. Ind. 66 (2015) 1–10.
[28] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733–1740.
[29] A. Zbrowski, Cleaning system for automated optical inspection of heads of bearing rollers, Solid State Phenom. 198 (2013) 289–294.
[30] F.-S. Yang, C.-C. Ho, L.-C. Chen, Automated optical inspection system for o-ring based on photometric stereo and machine vision, Appl. Sci. 11 (6) (2021) 2601.
[31] P. Banerjee, D. Chakravarty, B. Samanta, Design of a laboratory scale automatic optical inspection prototype system for scanning of conveyor belt surfaces-A case study, Measurement 220 (2023) 113342.
[32] N.H. Ali, R. Özdağ, Coverage analysis and a new metaheuristic approach using the Elfes probabilistic detection model in wireless sensor networks, Measurement 200 (2022) 111627.
[33] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: a natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[34] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[35] W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning, IEEE Trans. Neural Netw. Learn. Syst. 26 (6) (2014) 1275–1286.
[36] K.-Y. Lin, G. Wang, Hallucinated-IQA: no-reference image quality assessment via adversarial learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 732–741.
[37] W. Zhang, K. Ma, J. Yan, D. Deng, Z. Wang, Blind image quality assessment using a deep bilinear convolutional neural network, IEEE Trans. Circuits Syst. Video Technol. 30 (2019) 1.
[38] H. Zhu, L. Li, J. Wu, W. Dong, G. Shi, MetaIQA: deep meta-learning for no-reference image quality assessment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14143–14152.
[39] Z. Tang, Y. Chen, X. Chen, Xiaoping Zhang, Lightweight transformer and multi-head prediction network for no-reference image quality assessment, Neural Comput. Appl. 36 (4) (2024) 1947–1957.
[40] K. Xu, L. Liao, J. Xiao, C. Chen, H. Wu, Q. Yan, W. Lin, Boosting image quality assessment through efficient transformer adaptation with local feature enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 2662–2672.
[41] H. Talebi, P. Milanfar, NIMA: neural image assessment, IEEE Trans. Image Process. 27 (8) (2018) 3998–4011.
[42] J. Ke, Q. Wang, Y. Wang, P. Milanfar, F. Yang, MUSIQ: multi-scale image quality transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5148–5157.
[43] A. Mittal, R. Soundararajan, A.C. Bovik, Making a “completely blind” image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2012) 209–212.
[44] L. Chi, B. Jiang, Y. Mu, Fast fourier convolution, Adv. Neural Inf. Process. Syst. 33 (2020) 4479–4488.
[45] N. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, Appl. Comput. Harmon. Anal. 10 (3) (2001) 234–253.
[46] B. Patro, V. Agneeswaran, Scattering vision transformer: spectral mixing matters, Adv. Neural Inf. Process. Syst. 36 (2023) 54152–54166.
[47] D. Ghadiyaram, A.C. Bovik, Massive online crowdsourced study of subjective and objective picture quality, IEEE Trans. Image Process. 25 (1) (2015) 372–387.
[48] E. Prashnani, H. Cai, Y. Mostofi, S. Pradeep, PieAPP: perceptual image-error assessment through pairwise preference, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1808–1817.
[49] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, A. Bovik, From patches to pictures (PaQ-2-PiQ): mapping the perceptual space of picture quality, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3575–3585.
[50] G. Jinjin, C. Haoming, C. Haoyu, Y. Xiaoxing, J.S. Ren, D. Chao, PIPAL: a large-scale image quality assessment dataset for perceptual image restoration, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16, Springer, 2020, pp. 633–651.
[51] G. Wu, J. Jiang, K. Jiang, X. Liu, Fully 1× 1 convolutional network for lightweight image super-resolution, Mach. Intell. Res. 21 (2024) 1–15.
[52] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017) 5998–6008.
[53] I.O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, et al., MLP-Mixer: an all-MLP architecture for vision, Adv. Neural Inf. Process. Syst. 34 (2021) 24261–24272.
[54] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (1) (2017) 206–219.
[55] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, Y. Zhang, Blindly assess image quality in the wild guided by a self-adaptive hyper network, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3667–3676.
[56] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.