International Journal of Mechatronics and Applied Mechanics
2025, Issue 22, Vol. II

APPLICATION RESEARCH OF MULTI-SCALE FEATURE ATTENTION IMPROVEMENT ALGORITHM BASED ON RETINEX ALGORITHM AND DEEP LEARNING IN LOW-QUALITY IMAGE ENHANCEMENT

Tao Dong
Information Engineering School, Liaodong University, Dandong, 118000, China

Abstract - In recent years, intelligent driving technology has drawn attention to improving low-quality images in complex traffic environments. This study proposes a method combining multi-scale spatial attention mechanisms and the Retinex algorithm to enhance image detail and contrast. By integrating the Retinex algorithm with convolutional neural networks, the method optimizes brightness and sharpness. Additionally, a non-local attention module is used to capture remote dependencies and improve image semantics. Experimental results show that the proposed method achieves an SSIM of 0.96, a PSNR of 24.24, and reduces the running time by 83.12%, from 21.45 ms to 3.62 ms. The parameter count is also reduced from 25.06M to 20.45M, demonstrating significant computational efficiency. This method effectively enhances low-quality images, making it suitable for intelligent driving and real-time image processing applications.

Keywords: Multi-scale feature fusion, Attention, Image enhancement, Deep learning, Retinex algorithm.

1. Introduction

At present, with the acceleration of urbanization, the demand for monitoring and managing road traffic environment in various countries is increasing day by day. Images in road traffic environments are usually obtained through electronic cameras. Although existing imaging devices have to some extent solved the problem of traffic image degradation, these devices are expensive and bulky, making them difficult to popularize on a large scale. Even advanced imaging equipment cannot guarantee high-quality road traffic images under different weather conditions and complex traffic environments.

Therefore, the traffic images obtained by conventional electronic cameras often have problems such as color cast and low contrast [1]. In the context of accelerating urbanization, the intelligence and efficiency of traffic management are particularly important. Moreover, traditional electronic camera devices do face issues of decreased image quality, such as color cast and low contrast, in different weather conditions and complex traffic environments. Low-quality image enhancement has become a key research direction; therefore, the study of low-quality image enhancement has important practical significance. Thus, the main goal of the research is to improve the clarity and visual effects of low-quality images. In addition, due to multiple factors such as insufficient lighting, motion blur, and noise interference, low-quality images are commonly present in practical applications [2]. Recently, with the rapid growth of deep-learning, image enhancement algorithms grounded on convolutional neural networks have obtained certain success. However, traditional deep learning methods still face the contradiction of balancing noise suppression and detail enhancement when processing low-quality images [3]. Especially in complex real-world scenarios, single scale feature extraction often fails to fully capture multi-scale information in images, resulting in unsatisfactory enhancement effects. Introducing multi-scale features can enhance the optical effects of images by updating image quality, reducing noise, and improving details [4]. In addition, attention mechanism can enhance image clarity, reduce the loss of high-frequency details, and improve the clarity and contrast of low light images during the downsampling process of low-quality image enhancement. Retinex algorithm is a commonly used image enhancement algorithm that simulates the human visual system's perception ability of lighting and color, effectively enhancing the visibility and

https://doi.org/10.17683/ijomam/issue22.v2.16

visual contrast of images while preserving image details [5]. However, traditional Retinex algorithms suffer from high computational complexity and insufficient single scale feature extraction, which limits their effectiveness in complex traffic scenarios.

Therefore, this study innovatively utilizes multi-scale feature information, combines Retinex algorithm in an adaptive adjustment manner, and combines multi-scale feature extraction with attention mechanism for enhancing low-quality images, aiming to improve the overall clarity of images and optimize their visual effects. The primary contribution of this study lies in the integration of multi-scale feature extraction and attention mechanisms within the framework of Retinex-based image enhancement. By combining these techniques, the research significantly improves the performance of low-quality image enhancement, especially in challenging environments like low-light, hazy, or nighttime conditions.

The main objective of the research is to improve the quality of images in low light environments such as haze or nighttime. Under low light conditions, images often lack sufficient brightness and contrast, which makes it difficult to present details clearly and noise is easily noticeable. The Based on Multi Scale Feature Attention Improvement (MSFAI) algorithm proposed in the study, through multi-scale feature fusion and attention mechanism, can enhance image brightness while preserving important details and effectively suppress noise. The study chose PSNR and SSIM as evaluation metrics mainly because they can comprehensively reflect different aspects of image quality. PSNR is often used to measure the brightness and detail restoration ability of images, while SSIM can better evaluate the structural information and perceptual quality of images. These two indicators can simultaneously consider the brightness, contrast, and structural information of the image, providing a more comprehensive quality evaluation. Before conducting the experiment, the selected dataset was preprocessed. During the preprocessing process, the images were first standardized to ensure a uniform range of pixel values for all input images and eliminate bias caused by different image sources. Then, the low light images in the dataset were denoised to reduce the impact of noise on image quality evaluation.

The research content mainly includes five sections. The first part reviews the current research status on the role of image enhancement technology and deep learning technology in image enhancement both domestically and internationally. The second part explores low-quality image enhancement methods based on deep learning and provides a detailed introduction to an improved algorithm that combines multi-scale features and attention mechanisms. The third part validates the performance of the proposed image enhancement algorithm based on multi-scale feature attention improvement. The fourth part provides an in-depth analysis of the experimental results, exploring the reasons, advantages, generalization ability, and potential applications of the proposed algorithm in different fields to improve its performance. The last section summarizes the main achievements and core findings of the research, points out the limitations of the current study, and looks forward to future research directions.

2. Related Works

Image enhancement technology plays a key role in improving low-quality images and has therefore received attention from many researchers. Zhao Z et al. designed an image generation technique based on Retinex decomposition generation strategy to raise the contrast and detail clarity of low-light pictures. This technique utilized a united profundity structure to determine potential elements and enhance low-light pictures, and weakened the joint condition between the elements when performing Retinex putrefaction. The results indicated that this method had significant superiority [6]. Li C et al. [7] proposed a multi device dataset containing low-light pictures and videos to raise the quality of low-light pictures, and utilized a unified online platform to generate the results of picture improvement methods. The outcomes reflected that this way successfully improved the capacity of face detection in the dark. Wang Y et al. [8] raised a regularized flow pattern to effectively model the one to multi-mapping link between low-light pictures and normal-light pictures. This model employed a reversible network to ascertain the mapping of the distribution of normal-light pictures to a Gaussian distribution under low-light conditions, thereby facilitating a more accurate modelling of the conditional distribution of normal-light pictures. The results showed that this method significantly improved exposure illumination. Lv F et al. [9] proposed an end-to-end attention guidance system grounded on multi-branch convolutional neural networks to simultaneously address the issues of brightness restoration, fidelity deficiency, and noise in low-light picture improvement. The method employed two attention maps to facilitate the enhancement of brightness and the removal of noise, and the outcomes reflected that the system effectively generated high fidelity enhancement results.

Deep-learning plays a crucial role in image enhancement. In order to improve the visibility, color correction, and object detection performance in image enhancement under complex lighting environments, Liu R et al. [10] proposed using deep learning technology to construct a seabed image capture system, which used the object detection performance of enhanced images as a new evaluation criterion. The results showed that the system was effective. Zeng H et al. [11] raised a method of learning an image adaptive 3D lookup table to gain efficient and robust photo color and tone improvement. This method utilized a small convolutional neural network to predict weights related to image content, and integrated multiple basic 3D lookup tables into an adaptive enhanced lookup table. The outcomes showed that this way was more useful than the existing state-of-the-art photo enhancement methods. Shinozaki S et al. [12] proposed using deep learning techniques to compare the effectiveness of colorectal polyp detection in order to improve the quality of medical pictures.

Moreover, they used a random effects model to summarize the detection data of polyps and adenomas in the comparison. The results showed that this method effectively improved the detection rate of additional polyps in the right colon.

Hou Y et al. [13] proposed to use deep learning technology combined with invasive sensors to monitor the dynamic response structure of road surface in order to improve the monitoring effect of road traffic.
This method used image processing technology to monitor and evaluate the traffic road surface condition. The results showed that this method could effectively detect road surface damage. Zhou et al. [14] proposed an underwater image enhancement method using multi-interval sub-histogram perspective equalization to improve the clarity of underwater images.

This method estimated the degree of feature drift in each region of the image by extracting statistical features of the image, and used this information to guide feature enhancement for adaptive feature enhancement. The results showed that this method effectively improved the visual effect of degraded images. Guo, X. et al. [15] proposed a new framework inspired by the divide and conquer principle to solve the degradation problem of low light environment images. This framework decomposed images into texture and color components, and performed noise removal, color correction, and lighting adjustment.

The results showed that this method greatly reduced image degradation. The comparison table of different methods is shown in Table 1.

Table 1. Comparison Table of Different Methods

Method	PSNR	SSIM	Limitations
Zhao Z et al [6].	20.5	0.85	Limited contrast enhancement and insufficient detail recovery
Li C et al [7].	21.0	0.87	Dependent on multi device datasets, limited applicability
Wang Y et al [8].	00.2	0.90	High computational complexity and poor real-time performance
Lv F et al [9].	19.5	0.88	Sensitive to noise and slow processing speed
Liu R et al [10].	18.8	0.86	Limited applicability scenarios, algorithm optimization required
Zeng H et al [11].	20.5	0.91	Strong dependence on specific image content
Shinozaki S et al [12].	19.7	0.89	Mainly targeting medical images, with insufficient universality
Hou Y et al [13].	18.0	0.84	Monitoring accuracy is limited by sensor quality
Zhou et al [14].	22.1	0.92	Only applicable to underwater images, requiring strong prior information
Guo X et al[15].	23.0	0.93	Strong dependence on low light environments and high processing complexity

In summary, although there have been many achievements in using deep learning techniques for image enhancement, research on low-quality road traffic environments and combining multi-scale feature information with attention mechanisms for image enhancement is still very rare. Therefore, the application of multi-scale feature attention improvement algorithms in low-quality image enhancement is studied in order to improve the visual effect of low-quality images in road traffic environments.

3. Image Enhancement Algorithm Based on MSFAI

The research is conducted on improving the Retinex algorithm using the Hue, Saturation, Value (HSV) spatial model. In addition, to address limitations such as noise and color cast, further optimization is achieved by utilizing multi-scale features and attention mechanisms on the basis of improving the Retinex image enhancement structure.

3.1 Low-Quality Image Enhancement Based on Deep Learning Technology

Due to the influence of haze on light propagation, combined with low nighttime brightness and poor camera imaging quality, traffic road environment images often suffer from issues such as color deviation, image blurring, and even distortion. This results in difficulties in recognizing license plates, detecting traffic flow, and identifying accidents in these images, leading to a low quality of the images [16]. At present, traditional image enhancement algorithms have limited effectiveness in restoring low-quality images in diverse scenes, making it difficult to adapt to complex environments [17]. The Retinex algorithm in deep learning technology is a usually retrieved technique for picture improvement, whose core idea is to adjust the contrast and picture brightness while preserving the picture details [18]. Therefore, the research utilizes the network structure of Retinex based on deep learning technology as the foundation to design low-quality image enhancement. The network structure based on Retinex algorithm is shown in Figure 1.

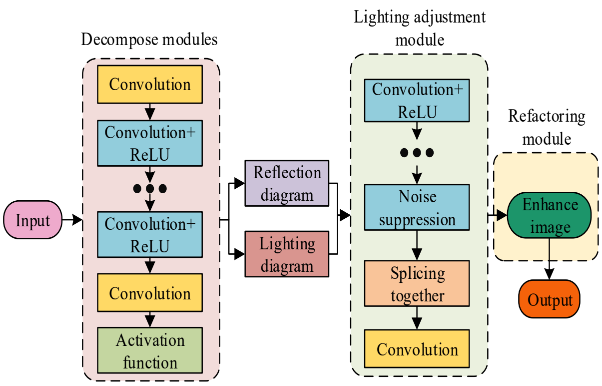

Figure 1: Network Structure Grounded on Retinex Algorithm

In Figure 1, the network structure grounded on Retinex algorithm mainly consists of three modules: decomposition, lighting adjustment, and reconstruction. Among them, the decomposition module performs convolution operation on the input image to separate the illumination and reflection parts in road traffic images under hazy light environment and low night brightness environment. The lighting adjustment module uses convolution and ReLU functions to suppress noise in the picture and adjust the lighting components of the picture. The reconstruction module recombines the decomposed reflection map and illumination map to generate an enhanced image. Throughout the process, the study utilizes a loss function for constraint to achieve adaptive adjustment of Retinex network structure to lighting.

The mathematical expression for reconstruction loss is shown in equation (1).

$$L_{Recon} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2 \tag{1}$$

In equation (1), L_{Recon} represents the reconstruction loss, n means the number of pixels in the input image, x_i means the i th feature of the original input data, and x_i' means the i th feature of the reconstructed data. The mathematical expression for reflection consistency loss is shown in equation (2).

$$L_{Ref} = \| R_{low} - R_{normal} \|_1 \tag{2}$$

In equation (2), L_{Ref} represents the reflection consistency loss, R_{low} means the reflection component of low-quality images, and R_{normal} means the reflection component of normal images. $\| R_{low} - R_{normal} \|_1$ represents the L1 norm of the reflection components of low-quality images and normal images. The math-expression for the smoothing loss of illumination is represented in equation (3).

$$L_s = \sum_{low,normal} \| \nabla I_m \cdot e^{-\lambda_g \nabla R_m} \| \tag{3}$$

In equation (3), L_s represents the smoothing loss of illumination, I_m means the illumination component, R_m means the light reflection component, λ_g means the intensity used to balance structural perception, and e represents the irrational base number. $\sum_{low,normal}$ represents summation, and ∇ represents gradient operation. The overall loss function of the Retinex network structure is shown in equation (4).

$$L_{Decom} = L_{Recon} + \lambda_r L_{Ref} + \lambda_s L_S \tag{4}$$

In equation (4), L_{Decom} represents the overall loss function, λ_s represents the illumination smoothing coefficient, and λ_r represents the balanced reflectance. Low-quality images of road traffic often have a strong contrast between brightness and background areas due to uneven lighting, which can lead to backlighting [19]. To effectively enhance such images, in addition to using the Retinex algorithm, it is also necessary to adapt the picture light appropriately.

Therefore, based on the Retinex network structure, the HSV color model and color spatial model are used to raise the adaptive adjustment of image brightness. The HSV spatial model is shown in Figure 2.

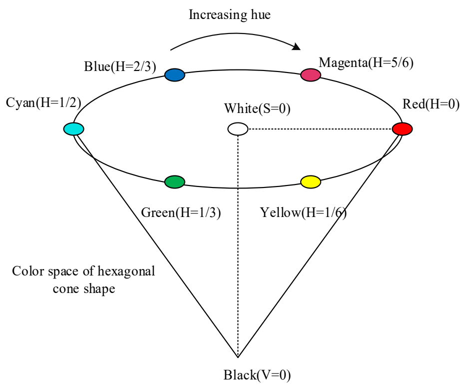

Figure 2: HSV Spatial Model

As shown in Figure 2, the HSV spatial model is a color space with a hexagonal pyramid shape, where H means the angle measure, S represents the saturation degree of the color close to the spectral color, and V represents the brightness of the color. This model can describe the color information perceived by the human eye, so the study uses the HSV spatial model to convert the Red, Green, Blue (RGB) in the picture to achieve adaptive brightness adjustment. The conversion expression of RGB color space is shown in equation (5).

$$H=\begin{cases} 0^\circ & max=min \\ 60^\circ\times\dfrac{G-B}{max-min}+0^\circ & max=R, G\geq B \\ 60^\circ\times\dfrac{G-B}{max-min}+360^\circ & max=R, G<B \\ 60^\circ\times\dfrac{B-R}{max-min}+120^\circ & max=G \\ 60^\circ\times\dfrac{R-G}{max-min}+240^\circ & max=B \end{cases}$$
$$S=\begin{cases} 0 & max=0 \\ \dfrac{max-min}{max} & other \end{cases}$$
$$V=max \tag{5}$$

In equation (5), H , S and V represent the conversion process from RGB to HSV color space, and R , B , and G respectively represent the coordinates of red, blue, and green in the color space. To separate the background of the image from the target features that need to be extracted, threshold segmentation is studied for handling the grayscale of the image. The mathematical expression for the average grayscale value is shown in equation (6).

$$mG=P_1\times m_1+P_2\times m_2 \tag{6}$$

In equation (6), mG represents the average grayscale value, P_1 and P_2 represent the proportion of target features and background pixels in the total image, and m_1 and m_2 represent the average grayscale values of target features and background pixels, respectively. The inter class variance between the target feature and the background is shown in equation (7).

$$\sigma^2=P_1P_2\left(m_1-m_2\right)^2 \tag{7}$$

In equation (7), σ^2 represents the inter class variance. The mathematical expression for brightness adjustment is shown in equation (8).

$$g(x,y,k,t)=\frac{-t}{\log\left(\dfrac{2kt+\frac{t}{50}}{\frac{t}{50}}\right)}\times\log\left(\frac{2kt+\frac{t}{50}-kx}{kx+\frac{t}{50}}\right)+t \tag{8}$$

In equation (8), g represents brightness adjustment, (x,y) represents the pixel position at the adjustment point, k represents the adjustment coefficient, and t represents the threshold. According to formula (8), brightness adjustment can be performed on dark areas in low-quality road traffic images based on the adjustment coefficient and threshold, so that the brightness of these areas can be moderately increased without distortion.

3.2 Low-Quality Picture Improvement Algorithm Based on MSFAI

The low-quality image enhancement network structure grounded on deep-learning has shown good performance in enhancing image brightness, but it has certain limitations in dealing with color cast, noise, and other issues in low-light images. Therefore, further research is needed to improve the Retinex image enhancement structure by combining multi-scale feature extraction and attention mechanisms for optimization. The attention mechanism originates from the study of the biological visual system, and its core idea is to focus on the regions of interest and ignore irrelevant regions. In deep learning, attention mechanisms enable networks to filter input features and assign different weights to different features, thereby improving the expressive power of key features.

Therefore, the study utilizes the non-local interest module in self attention mechanism to consider the relationship between each feature point and all other feature points, aiming to capture long-range dependencies. The basic logic of the non-local interest module is represented in Figure 3.

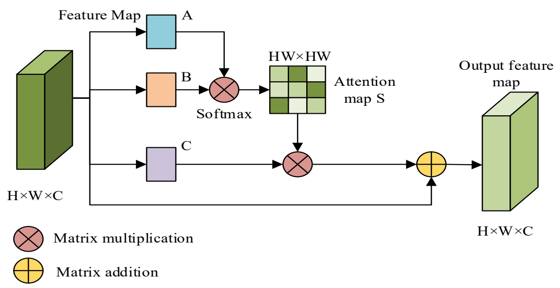

Figure 3: Non-Local Attention Module Structure

In Figure 3, the non-local interest module generates three different feature maps A, B, and C by linearly transforming the input image. Then, Softmax normalization on feature maps A and B is performed to generate attention map S. Next, the attention map S is matrix multiplied with the feature map C and summed with the initial input image to obtain the output feature map that captures remote dependencies. The weight expression of the attention mechanism is shown in equation (9).

$$y_\alpha = \sum_{\beta=1}^{N} \frac{f(x_\alpha, x_\beta)}{C(x)} \cdot \psi(x_\alpha) \tag{9}$$

In equation (9), y_α means the weight of the feature located at position α, $f(x_\alpha, x_\beta)$ represents the correlation function between positions α and β, N means the total number of features, C means normalization, and ψ represents linear mapping. Multi-scale feature technology can extract image features at various spatial scales, thereby enhancing the semantic information of low-quality images. The process based on MSFAI is shown in Figure 4.

In Figure 4, in the MSFAI process, the input image is first subjected to feature extraction through a convolutional neural network, and the relationship between image features is captured through a non-local attention module. Subsequently, multi-scale feature extraction techniques are used to improve the details of the image, and concatenation is performed grounded on the relationships between features to aggregate multi-scale detail information. Finally, it is processed through deconvolution and channel equalization attention modules to achieve the goal of optimizing image resolution. The feature extraction for convolution kernels of different scales is shown in equation (10).

$$f_1 = Concat[relu(Conv_{3\times3}(x_{in})), relu(Conv_{5\times5}(x_{in}))] \tag{10}$$

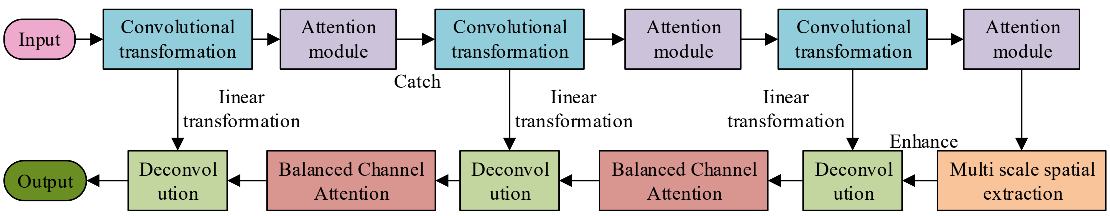

Figure 4: Improving the Process Based on MSFAI

Equation (10), f_1 represents the output feature matrix of the first layer spatial convolution kernel, $Concat$ represents concatenation operation, $relu$ means activation function, $Conv$ means convolution operation, and x_{in} represents input feature matrix. The feature extraction for the second layer spatial scale convolution kernel is shown in equation (11).

$$f_2 = relu\left(Conv_{3\times3}\left(relu(Conv_{3\times3}(f_1))\right)\right) + f_1 \tag{11}$$

In equation (11), f_2 represents the output feature matrix of the second layer spatial convolution kernel. By combining the operations of equations (10) and (11), the feature extraction for multi-layer spatial scale convolution kernels can be obtained as shown in equation (12).

$$f_k = relu\left(Conv_{1\times1}\left(relu\left(Conv_{3\times3}(f_{k-1})\right)\right)\right) + x_{in} \tag{12}$$

In equation (12), f_k represents the output feature matrix of the k th layer spatial convolution kernel. Through the feature extraction and concatenation operations of multi-layer convolution, multi-scale features can be fused to achieve hierarchical connections at multi-scales. The mathematical expression of the weight coefficient is shown in equation (13).

$$W_C = sigmoid\left(MLP(AvgPool(x_{in}))\right) \tag{13}$$

In equation (13), W_C represents the weight coefficient, MLP represents the shared network, and $AvgPool$ represents average pooling.

The math-expression of the global similarity loss function is shown in equation (14).

$$L_1(G) = E[\|1 - G(X,Z)\|_1] \tag{14}$$

In equation (14), L_1 represents the global similarity loss function, G represents the generator, E represents the expectation, X means the real image data, and Z means noise.

The mathematical expression of the content aware loss function is shown in equation (15).

$$L_{con}(G) = E[\|1 - \Phi(G(X,Z))\|_2] \tag{15}$$

In equation (15), L_{con} represents the content aware loss function.

The architecture based on MSFAI algorithm is shown in Figure 5.

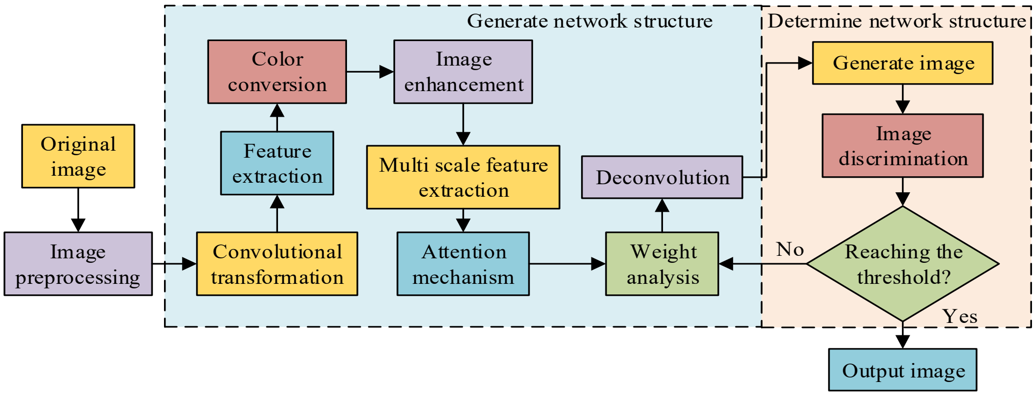

Figure 5: Architecture Based on MSFAI Algorithm

As shown in Figure 5, the architecture based on MSFAI algorithm can be mainly divided into two modules: generating network structure and determining network structure. In the generation network module, convolution transformation and other operations are used to enhance image features, followed by color conversion through the HSV space model, and multi-scale feature extraction techniques are applied to improve the semantic data of the image. Afterwards, the attention structure is retrieved for weight analysis, and finally the enhanced image is generated through deconvolution processing. The judgment network module is responsible for evaluating the enhancement effect, and it outputs the final image when the judgment image reaches the preset enhancement threshold. If the effect is not achieved, it will return to the generation network module for weight analysis and optimization again.

4. Validation of Image Enhancement Algorithm based on MSFAI

The study first establishes an experimental environment, then validates the image enhancement algorithm and network model complexity, and finally conducts practical application verification.

4.1 Experimental Environment Construction

To testify the capacity of the MSFAI-based image improvement algorithm, an experimental environment is first set up. The algorithm is conducted in the PyTorch structure, and the hardware environment used is Windows 10 operating system. The processor is Xeon (R) Silver 4114 CPU, equipped with Tesla V100 GPU, and configured with 32GB of video memory and memory. The datasets used in the experiment are the KITTI dataset and the Cityscapes dataset. The KITTI dataset contains a large number of road traffic scene images, including vehicles, pedestrians, and traffic signs, while the Cityscapes dataset contains rich urban traffic scene images for tasks such as semantic segmentation and object detection. The study divides these two datasets into training and testing sets in a 3:7 ratio. The specific parameters are represented in Table 2.

Table 2. Specific Experimental Environment

Experiment Environment	Configuration
Operating System	Windows 10
CPU	Xeon(R) Silver 4114
GPU	Tesla V100
Memory	32G
Video Memory	32G
Framework	PyTorch

4.2 Performance Verification of Picture Improvement Algorithms

To verify the capacity of the MSFAI-based picture improvement algorithm, a comparative analysis is conducted between this algorithm and other advanced image enhancement algorithms. The experiments conducted on the KITTI and Cityscapes datasets both used the same hyperparameter settings. In formula (13), the selection of weight coefficients is based on experimental experience and parameter tuning results. In order to ensure the optimal performance of MSFAI algorithm in image enhancement, an optimization method based on gradient descent was adopted to adjust these weight coefficients. Adam optimizer was used in the optimization process research, and an early stop strategy was adopted to prevent overfitting.

The specific hyperparameter settings are as follows: learning rate of 0.04, batch size of 16, weight decay of 0.05, and training wheel speed of 50 rounds. The software versions are as follows: Python version is 3.8.10, PyTorch version is 1.10.0, and CUDA version is 11.3. The code repository is https://github.com/your-repository. The compared algorithms include Representation Forgetting Unlearning (RFU) algorithm [20], Underwater Convolutional Neural Network (UWCNN) algorithm [21], and traditional Retinex algorithm [22]. The study uses the No-Reference Image Quality Metric for Contrast Distortion (NIQMC) value to estimate the impact of different algorithms on picture contrast distortion quality. The comparison of image enhancement performance of different algorithms is shown in Figure 6.

From Figure 6 (a), the KITTI dataset, the NIQMC value of the MSFAI algorithm consistently outperformed other algorithms, with a max NIQMC value of 5.35. Compared with the maximum NIQMC values of 4.58, 4.25, and 2.84 for RFU, UWCNN, and Retinex algorithms, they increased by 14.39%, 20.56%, and 46.91%, respectively. From Figure 6 (b), in the Cityscapes dataset, the maximum NIQMC value of the MSFAI algorithm was 5.47.

Compared to the maximum NIQMC values of 4.50, 3.97, and 3.05 for RFU, UWCNN, and Retinex algorithms, the MSFAI algorithm improved by 17.73%, 27.42%, and 44.24%, respectively. In summary, the MSFAI algorithm exhibited superior image enhancement performance on both KITTI and Cityscapes datasets, significantly improving the contrast quality of road traffic environment images.

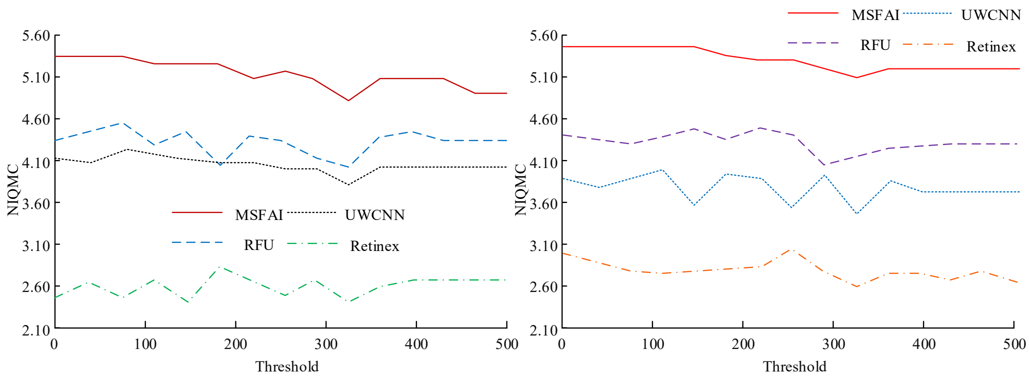

Figure 6: Comparison of Picture Improvement Performance of Different Algorithms

Table 3. Comparison Results of Image Quality Indicators for Different Algorithms

Evaluating indicator	Data set	MSFAI	RFU	UWCNN	Retinex
RTIQM	KITTI	3.08	2.65	2.81	2.77
	Cityscapes	3.06	2.51	2.83	2.75
RTCIQE	KITTI	9.75	2.83	4.50	3.85
	Cityscapes	9.51	2.77	5.03	3.92
PSNR	KITTI	22.01	15.03	18.32	17.06
	Cityscapes	21.97	14.84	18.49	17.14

In order to further validate the performance of the MSFAI image enhancement algorithm, multiple image quality evaluation metrics were used for comparative analysis, including Road Traffic Image Quality Metric (RTIQM), Road Traffic Color Image Quality Evaluation (RTCIQE), and Peak Signal to Noise Ratio (PSNR). The comparison results of image quality indicators for different algorithms are shown in Table 3. From Table 3, the MSFAI algorithm significantly outperformed other algorithms in various image quality evaluation metrics across different datasets. Among them, the maximum RTIQM value of MSFAI algorithm reached 3.08, which was 13.96%, 8.76%, and 10.06% higher than RFU, UWCNN, and Retinex algorithms, respectively. The maximum RTCIQE value of MSFAI algorithm was 9.75, which was 70.97%, 53.84%, and 60.51% higher than other algorithms, respectively. The maximum

PSNR value was 22.01, which was 31.71%, 16.76%, and 22.48% higher than other algorithms, respectively. Overall, the MSFAI algorithm had higher effectiveness and robustness in improving image quality.

The Structural Similarity (SSIM) metric can estimate image similarity from three angles: brightness, contrast, and structure. Therefore, the SSIM index was retrieved to esetimate the similarity between the enhanced image and the input image. The comparison of the enhancement effects of different algorithms on low-quality road traffic environment images is represented in Figure 7. From Figure 7 (a), in the KITTI dataset, the SSIM value of the MSFAI algorithm was 0.96, while the SSIM values of the RFU, UWCNN, and Retinex algorithms were 0.86, 0.68, and 0.80, respectively. The MSFAI algorithm improved by 10.41%, 29.16%, and 16.66%, respectively. From Figure 7 (b), in the Cityscapes dataset, the SSIM value of the MSFAI algorithm was 0.92, which was 25.00%, 47.82%, and 27.17% higher than the 0.69, 0.48, and 0.67 values of the other three algorithms, respectively. From a qualitative analysis perspective, the research method on the KITTI dataset outperforms other methods in enhancing low light images. It can be observed that the MSFAI algorithm performs particularly well in enhancing the overall brightness, detail restoration, and noise suppression of images.

Especially in the processing of dark details and background noise in images, MSFAI can significantly improve the clarity of images, making details more abundant, while other algorithms often result in loss of details or amplification of noise while increasing brightness. On the Cityscapes dataset, the research method not only improved the brightness of low light images, but also showed a more balanced performance in contrast and detail restoration.

The research method restored the road details that were difficult to identify in the original image, and with proper noise control, the image quality was significantly improved. From the above, it is proved that the low-quality picture improvement effect based on MSFAI algorithm performed well in terms of brightness, contrast, and structure, which can significantly improve the quality of the image and make it closer to a high-quality reference image.

(a) Image Enhancement Effect on KITTI Dataset

(b) Image Enhancement Effect on Cityscapes Dataset

Figure 7: Comparison of road traffic environment image enhancement effects using different algorithms

4.3 Verification of Network Model Complexity

To verify the complexity of the improved algorithm model based on MSFAI, centralized testing was conducted on the model structures of different algorithms in the test set, and the running time and parameter count of each algorithm were compared. The comparison results of the running time and parameter quantity of different algorithms are shown in Figure 8. From Figure 8 (a), the average operating duration of the MSFAI algorithm was 3.62ms, which was only 1.91ms longer than the fastest UWCNN algorithm, although not the shortest compared to other algorithms. In contrast, the operating duration of Retinex and RFU algorithms were 21.45ms and 27.70ms, respectively. Overall, the MSFAI algorithm reduced the running time by 83.12% and 86.93% compared to Retinex and RFU algorithms, respectively. From this, the MSFAI algorithm not only enhanced network performance, but also demonstrated significant advantages in computational speed. From Figure 8 (b), the parameter count of the MSFAI algorithm was 20.45M. Compared with the parameter sizes of 14.87M, 19.72M, and 25.06M for RFU, UWCNN, and Retinex algorithms, the parameter size of MSFAI algorithm only increased by 5.58M and 0.73M compared to RFU and UWCNN. In addition, the PSNR value of the MSFAI algorithm was 24.24, significantly higher than the other three algorithms. Compared with the PSNR values of RFU, UWCNN, and Retinex algorithms, the PSNR of MSFAI algorithm increased by 38.03%, 28.91%, and 32.71%, respectively. The MSFAI algorithm still maintained a significant PSNR advantage with a slight increase in parameter count.

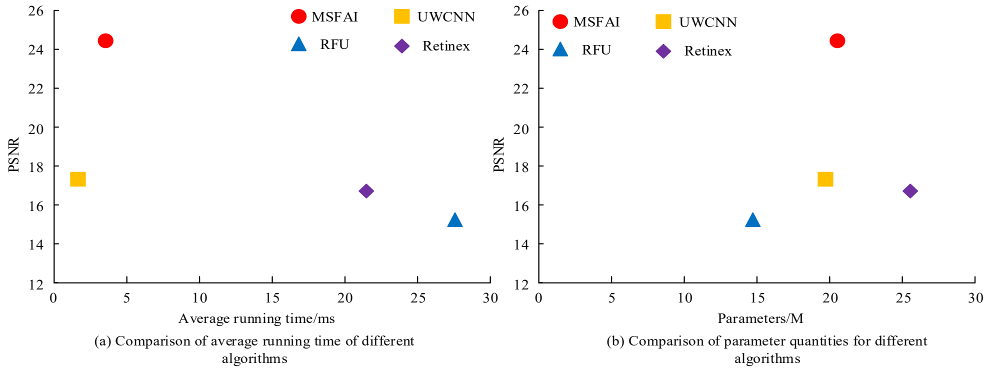

Figure 8: Comparison Results of Running Time and Parameter Quantity of Different Algorithms

To make further efforts on testifying the capability of each module in the MSFAI algorithm, the research conducted ablation experiments to test each module. The experiment used the Blind Tone Mapped Quality Index (BTMQI) as the evaluation criterion, which is used to evaluate the image quality after tone mapping processing and can effectively reflect the improvement effect of image enhancement technology. The ablation experiment is represented in Table 4. From Table 4, it is crystal that when using only the Retinex network structure, the BTMQI value was 2.97. When using HSV for improvement, the BTMQI value increased to 3.20, with an increase of 7.74%. After adding attention mechanism, the BTMQI value further increased to 3.42, with an increase of 6.87%. When multi-scale features were introduced, the BTMQI value significantly increased to 4.33, with an increase of 26.60%. After adding attention mechanism, the BTMQI value was increased to 3.42, which was a 15.15% improvement compared to the benchmark model that only used Retinex network structure. After adding multi-scale features, the BTMQI value of the model was increased to 4.33, which was 45.79% higher than the baseline model. It can be seen that the addition of each module effectively improves the

Table 4. Ablation Experiment

Case	Retinex	HSV	Attention module	Multi-scale features	BTMQI
1	√	/	/	/	2.97
2	√	√	/	/	3.20
3	√	√	√	/	3.42
4	√	√	√	√	4.33

4.4 Practical Application Verification

To testify the capacity of the MSFAI algorithm in practical applications, the study compares the subjective evaluations of different algorithms using the Average Subjective Opinion Score (MOS), and a significance test was conducted on it. The study randomly selected 10 male and female users and subjectively scores 20 groups of enhanced images. The MOS statistical results are represented in Figure 9. From Figure 9, the mean MOS of the MSFAI algorithm was 3.96 with a standard deflection of 0.42, while the mean and deviation of the RFU algorithm were 3.15 and 0.71, respectively, the statistical results showed significant differences ($P<0.05$). The mean and deviation of the UWCNN algorithm were 1.85 and 0.74, respectively, while the mean and deviation of the Retinex algorithm were 2.96 and 0.25, the statistical results showed significant differences ($P<0.05$). From this, it is proved that the picture improvement effect of MSFAI algorithm in practical applications can better meet the optical requirements of the human eyesight.

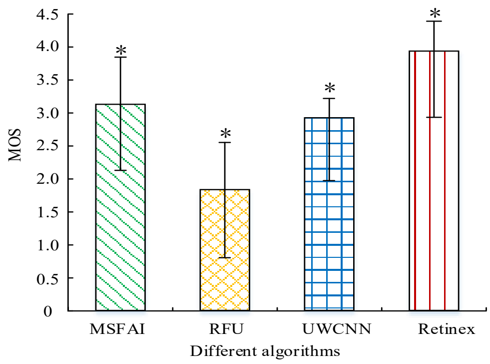

Figure 9: MOS Statistical Results (indicates P<0.05)*

5. Discussion

In the study, MSFAI was compared with several state-of-the-art image enhancement methods, and the results showed that MSFAI exhibited significant improvements in both PSNR and SSIM metrics. For example, the PSNR value of the MSFAI-improved algorithm was 24.24, significantly higher than the other three algorithms. Compared with the PSNR values of RFU, UWCNN, and Retinex algorithms, the PSNR of MSFAI algorithm increased by 38.03%, 28.91%, and 32.71%, respectively. Compared with other advanced algorithms, the PSNR and SSIM of the research method also had advantages. For example, in the KITTI dataset, the SSIM value of the MSFAI algorithm was 0.96 and the PSNR value was 24.24. The SSIM and PSNR of the study were 0.85 and 20.5, respectively, and the research method improved by 12.94% and 18.24% compared to them. This is mainly due to the combination of MSFAI and attention mechanism proposed in the research. Multi-scale feature fusion enables algorithms to better process image information at different scales, preserving more details and textures, while attention mechanisms effectively suppress background noise when enhancing important areas of the image, further improving image quality.

The reason behind the enhanced performance of the research methods lies in the multi-scale feature fusion, which effectively boosts the algorithm's capability to capture image details by processing image features across various scales. Especially in low-light environments, this technique can efficiently restore dark details and mitigate overexposure of bright areas. Secondly, the attention mechanism assigns higher weights to important regions in the image, further enhancing the clarity of the image, especially in complex backgrounds. It can highlight the core information in the image and suppress interference, improving the overall quality of the image. The MSFAI algorithm not only performs excellently in standard low light image enhancement tasks, but also has strong practical application value. For example, in underwater image enhancement, MSFAI can effectively remove underwater noise, restore image details, and provide clearer and more realistic underwater visual effects than traditional methods. In addition, the study also tested the performance of MSFAI in hazy weather and nighttime images. The experimental results showed that MSFAI could better restore image details and contrast under these low light or blurry conditions. Compared with traditional methods, MSFAI has significant advantages in image clarity and detail preservation. Therefore, MSFAI can be widely applied in fields such as intelligent monitoring, autonomous driving, medical imaging, etc., especially in image processing tasks in low light or complex environments with enormous potential.

Although research mainly focuses on enhancing underwater images, the results of MSFAI algorithm show strong generalization ability. Therefore, by combining multi-scale feature fusion and attention mechanisms, MSFAI can adapt to various low light environmental conditions such as fog or nighttime images.

6. Conclusions

The image quality under complex lighting conditions significantly affects the accuracy and reliability of visual tasks. To raise the definition and clarity of low-quality images, the Retinex algorithm was optimized using multi-scale feature extraction combined with attention mechanism to enhance its ability to capture image details. The results showed that in the KITTI dataset, the max NIQMC value of the MSFAI algorithm reached 5.35, which was 14.39%, 20.56%, and 46.91% higher than the max NIQMC values of the RFU, UWCNN, and Retinex algorithms, respectively. In the Cityscapes dataset, the max NIQMC value of the MSFAI algorithm was 5.47, which was 17.73%, 27.42%, and 44.24% higher than the max NIQMC values of other algorithms, respectively. The MSFAI algorithm significantly outperformed algorithms such as RFU, UWCNN, and Retinex in terms of image quality indicators such as PSNR and SSIM. Among them, the PSNR and SSIM values of MSFAI were 24.24 and 0.96, respectively, which were significantly better than other algorithms. In addition, the capacity of MSFAI algorithm in terms of running time and parameter quantity was also relatively balanced, with 3.62ms and 20.45M, respectively. In the ablation experiment, when multi-scale features were introduced, the BTMQI value significantly increased to 4.33, with an increase of up to 26.60%. In summary, the application of MSFAI algorithms in low-quality image enhancement had effectively improved the visibility and clarity of low-quality images. In the field of medical imaging, the MSFAI algorithm can be used to enhance low contrast or blurry medical images, helping to improve the accuracy of disease diagnosis. In the field of security monitoring, this algorithm can be used to optimize monitoring images under low light or adverse weather conditions, improving the effectiveness of security monitoring systems. However, the study only analyzed low-quality images of road traffic environments. With the diversification of application requirements, future research should further focus on optimization under extreme lighting conditions.

For example, in extremely strong or low light environments, image quality may significantly decrease, and existing image enhancement methods may not be able to effectively restore details. Therefore, future research can improve the robustness and adaptability of algorithms under these extreme lighting conditions by introducing more complex lighting models or deep learning techniques, thereby expanding the application scope of algorithms.

Acknowledgement

The research is supported by: University Basic Scientific Research Project of Liaoning Provincial Department of Education in the year of 2023: Research on lightweight Road Environment perception and Driving area Detection (Fund No.: JYTMS20230709); Chinese Universities' New Generation Information Technology Innovation Projectin in the year of 2024: Research on Road Environment and Traffic Space Detection for Autonomous Driving (Fund No.: 2023IT188). Liaoning Provincial Science and Technology Program Joint Program in the year of 2024: Research on key technologies of road target detection under rainy and foggy environment (Fund No.: 2024JH2/102600144).

References

[1] Hao, S., Han, X., Guo, Y., Xu, X., & Wang, M. (2020). Low-light image enhancement with semi-decoupled decomposition. IEEE Transactions on Multimedia, 22(12), 3025–3038. https://doi.org/10.1109/TMM.2020.2969790

[2] Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., ... Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. Nature Communications, 11(1), 2624–2631. https://doi.org/10.1038/s41467-020-15948-9

[3] Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5), 820–838. https://doi.org/10.1109/JPROC.2021.3054390

[4] Guo, X., Yang, J., & Yang, L. (2024). Retrieval and analysis of multimedia data of robot deep neural network based on deep learning and information fusion. Informatica, 48(13). https://doi.org/10.31449/inf.v48i13.6063

[5] Zhao, B., Wang, C., Fu, Q., & Han, Z. (2020). A novel pattern for infrared small target detection with generative adversarial network. IEEE Transactions on Geoscience and Remote Sensing, 59(5), 4481–4492. https://doi.org/10.1109/TGRS.2020.3012985

[6] Zhao, Z., Wang, L., Ou, Q., Yu, L., & Kuang, F. (2021). RetinexDIP: A unified deep framework for low-light image enhancement. IEEE Transactions on Circuits and Systems for Video Technology, 32(3), 1076–1088. https://doi.org/10.1109/TCSVT.2021.3073371

[7] Li, C., Guo, C., Han, L., Jiang, J., Cheng, M. M., Gu, J., & Loy, C. C. (2021). Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12), 9396–9416. https://doi.org/10.1109/TPAMI.2021.3126387

[8] Wang, Y., Wan, R., Yang, W., Li, H., Chau, L. P., & Kot, A. (2022). Low-light image enhancement with normalizing flow. Proceedings of the AAAI Conference on Artificial Intelligence, 36(3), 2604–2612. https://doi.org/10.1609/aaai.v36i3.20162

[9] Lv, F., Li, Y., & Lu, F. (2021). Attention guided low-light image enhancement with a large scale low-light simulation dataset. International Journal of Computer Vision, 129(7), 2175–2193. https://doi.org/10.1007/s11263-021-01466-8

[10] Liu, R., Fan, X., Zhu, M., Hou, M., & Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology, 30(12), 4861–4875. https://doi.org/10.1109/TCSVT.2019.2963772

[11] Zeng, H., Cai, J., Li, L., Cao, Z., & Zhang, L. (2020). Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(4), 2058–2073. https://doi.org/10.1109/TPAMI.2020.3026740

[12] Shinozaki, S., Kobayashi, Y., Hayashi, Y., Sakamoto, H., Sunada, K., Lefor, A. K., & Yamamoto, H. (2020). Colon polyp detection using linked color imaging compared to white light imaging: Systematic review and meta-analysis. Digestive Endoscopy, 32(6), 874–881. https://doi.org/10.1111/den.13613

[13] Hou, Y., Li, Q., Zhang, C., Lu, G., Ye, Z., Chen, Y., Wang, L., & Cao, D. (2021). The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis. Engineering, 7(6), 845–856. https://doi.org/10.1016/j.eng.2020.07.030

[14] Zhou, J., Pang, L., Zhang, D., & Zhang, W. (2023). Underwater image enhancement method via multi-interval subhistogram perspective equalization. IEEE Journal of Oceanic Engineering, 48(2), 474–488. https://doi.org/10.1109/JOE.2022.3223733
[15] Guo, X., & Hu, Q. (2023). Low-light image enhancement via breaking down the darkness. International Journal of Computer Vision, 131(1), 48–66. https://doi.org/10.1007/s11263-022-01667-9
[16] Jian, Y., Dong, X., & Jian, L. (2021). Detection and recognition of abnormal data caused by network intrusion using deep learning. Informatica, 45(3). https://doi.org/10.31449/inf.v45i3.3639
[17] Chan, J. Y. L., Bea, K. T., Leow, S. M. H., Phoong, S. W., & Cheng, W. K. (2023). State of the art: A review of sentiment analysis based on sequential transfer learning. Artificial Intelligence Review, 56(1), 749–780. https://doi.org/10.1007/s10462-022-10183-8
[18] Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., & Saria, S. (2021). The clinician and dataset shift in artificial intelligence. New England Journal of Medicine, 385(3), 283–286. https://doi.org/10.1056/NEJMc2104626
[19] Adamkiewicz, M., Chen, T., Caccavale, A., Gardner, R., Culbertson, P., Bohg, J., & Schwager, M. (2022). Vision-only robot navigation in a neural radiance world. IEEE Robotics and Automation Letters, 7(2), 4606–4613. https://doi.org/10.1109/LRA.2022.3150497
[20] Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering, 42(1), 79–107. https://doi.org/10.1016/j.bbe.2021.11.004
[21] Andriyani, W. (2023). Influence of product quality and brand image on purchasing decisions for ABC Ground coffee products in South Tangerang. Digital Business: Tren Bisnis Masa Depan, 14(2), 67–74. https://doi.org/10.59651/digital.v14i2.109
[22] Tosaki, T., Yamakawa, M., & Shiina, T. (2023). A study on the optimal condition of ground truth area for liver tumor detection in ultrasound images using deep learning. Journal of Medical Ultrasonics, 50(2), 167–176. https://doi.org/10.1007/s10396-023-01301-2