Dual-Channel Convolutional Neural Network with Mask Guidance for Digital Media Image Super-Resolution

Qian Song
Department of Animation Arts, Zibo Vocational Institute, Zibo 255300, China
E-mail: zibosongqian@163.com

Keywords: image design, reconstruction, digital media, image resolution, convolutional neural network, super resolution

Received: July 18, 2025

To address the issues of low resolution and detail loss in digital media visual image reconstruction, this paper proposes a super-resolution reconstruction algorithm based on a dual-channel mask-guided convolutional neural network (CNN). This algorithm innovatively designs a parallel guidance channel that utilizes edge masks generated from the input image as prior knowledge to enhance the main channel's ability to extract high-frequency structural features. Through multi-layer feature fusion and improved Adam optimizer, the model achieves a balance between reconstruction quality and efficiency. Tested on the CIFAR-10 and ILSVRC2020 datasets, and validated using image quality assessment metrics, it was found that under super-resolution tasks, the peak signal-to-noise ratio of the research method improved by an average of about 1.1 dB, with a maximum of 37.12 dB. The structural similarity index improved by an average of about 2.1%, with a maximum of 0.9682. At the same time, the algorithm demonstrated excellent efficiency advantages while maintaining high performance, with model parameters reduced to 1.2M, which was only 1/35 of advanced models in the same category. The average inference time for a single image has been reduced to about 20 ms, which was about 45% faster than the baseline method and demonstrated superior overall performance. The improved algorithm offers improved reconstruction quality and achieves a balance between quality and efficiency when the number of parameters is limited. Improving CNN through dual channel collaborative design and feature reuse mechanisms can provide new technological solutions for digital media image processing.

Povzetek: Predlagan je učinkovit dvo-kanalni CNN algoritem za super-ločljivost slik, ki z uporabo robnih mask izboljša kakovost rekonstrukcije ob hkratnem zmanjšanju števila parametrov in časa izvajanja.

1 Introduction

In recent years, digital media (DM) has emerged as a prominent focus of research due to its increasing applications in communication, visualization, and technology-driven solutions. The enhancement and reconstruction of visual images in the DM domain are key research areas, with a focus on improving image quality, resolution, and interpretability. The purpose is to develop methodologies for designing and reconstructing DM visuals to achieve higher realism and fidelity [1]. With advancements in relevant technology, image generation (IG) has found extensive applications in computer vision, including virtual reality (VR), augmented reality (AR), and automated image synthesis [2]. At present, there are many methods for designing and reconstructing DM images. However, high computational complexity may occur during the operation of these methods. This is mainly due to extensive parameter tuning and the need for large datasets. These issues lead to longer training times, and issues such as low image resolution often result in suboptimal reconstruction quality, limiting the practicality of these methods [3]. The development of current visual interactive images is hindered by factors such as low resolution, dynamic lighting changes, occlusion, and scale variations. Traditional super-resolution reconstruction algorithms primarily rely on either frequency domain or spatial domain features, with limited integration of both [4-5]. The current DM visual image reconstruction faces three major technological bottlenecks: 1. Insufficient feature utilization: Traditional super-resolution algorithms, such as interpolation and sparse coding, rely on artificially designed features and are difficult to adaptively extract multi-scale semantic information (such as edges and textures) from images. This results in problems such as edge blurring and loss of detail that often appear in the reconstruction results. 2. Conventional machine learning-based super-resolution reconstruction algorithms have disadvantages such as large data volume, complex computation, and inability to utilize some prior features of images, making it difficult to meet real-time requirements. 3. Lack of prior knowledge: The existing methods do not fully utilize the inherent features of the image, such as occluded edges, resulting in insufficient robustness to occlusion and dynamic lighting [6-7].

In response to the above challenges, various deep learning architectures have emerged in recent years. U-Net achieves multi-scale feature fusion through encoder-decoder structure and skip connections, and performs excellently in image reconstruction tasks [8]. Generative adversarial networks (GANs) enhance visual realism through adversarial training and are widely used in perception-oriented super-resolution reconstruction [9]. However, the complex structure of U-Net may result in high computational overhead, while the training of GAN is unstable and prone to introducing artifacts. In addition, existing methods rely heavily on network self-learning features and lack explicit guidance on image structure priors, which limits the accuracy in detail restoration and noise balance. This study aims to solve the above technical bottlenecks and meet the urgent demand for high-fidelity image reconstruction technology in digital imaging applications such as VR/AR and intelligent film and television production. It attempts to improve the quality of image feature fusion and reconstruction, and release the potential of hybrid modeling technology that combines data-driven and existing knowledge. Therefore, on the basis of existing research, this paper takes advantage of the super-resolution convolutional neural network (SRCNN) and proposes an improved CNN image super-resolution reconstruction algorithm under dual-channel-guided convolution. Moreover, guidance channels and mask images are introduced to enhance feature extraction capabilities. The innovation of this study is reflected in two aspects: First, the mask-guided dual-channel architecture. That is, based on the classic SRCNN, a second channel is introduced to specifically input the mask image generated by the residual transformation. It explicitly enhances high-frequency detail reconstruction through an adaptive feature fusion mechanism and prior edges to achieve a balance between noise reduction and detail preservation. The second is a lightweight design architecture. That is, the local structure is enhanced through the guidance channel, avoiding complex encoding and decoding or adversarial training, and improving the reconstruction efficiency and stability while maintaining a low number of parameters.

The research content is mainly divided into four sections. The subsequent section is organized as the literature section, which reviews the classic methods of image super-resolution reconstruction and related deep learning methods. The third section introduces the SRCNN reconstruction method and improvement ideas, including the addition of a guidance channel. The fourth section introduces the experimental design details and reconstruction performance of the improved CNN algorithm. The final section introduces the conclusions and future scope of the current research, as well as research prospects.

2 Related works

DM has gained traction in recent years due to its wide range of applications in communication, entertainment, and computational imaging. Wang et al. proposed a multi-scale extended CNN framework for image compression sensing and reconstruction by jointly training a fully convolutional structure measurement network and a reconstruction network. The results show that it outperforms existing methods in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [10]. Chen et al. proposed a two-stage image restoration (IR) network based on a parallel network and contextual attention to solve problems such as blurring and texture distortion in infrared. The results showed that this method could achieve a more realistic reduction effect [11]. Nakamura T utilized wave-optical deconvolution filters and color channel image synthesis to refine image reconstruction. The results showed that the research method overcame the traditional limitations and effectively improved the image resolution [12]. Pan X et al. proposed a transformer-based fully connected neural network for mask-based frameless image reconstruction, which improves the reconstruction ability through global feature inference. The results showed that it was superior to model-based and fully convolutional network methods [13]. Low-resolution images often resulted in loss of fine-grained details due to insufficient sampling rate and compression artifacts, and Chen et al. suggested that the low-frequency and high-frequency components of feature maps were treated equally in existing image super-resolution reconstruction methods [14]. Low-resolution images were difficult to accurately depict complex surface structures, and the complexity of spatial variations and structural features made it difficult to restore image structures, especially in some photometric stereoscopic images [15]. Yang et al. believed that sparse structural similarity could be used to evaluate the quality of distorted images, thereby improving the evaluation of image quality by ignoring the shortcomings of image structure. The results showed that the Pearson correlation coefficient of this method was 0.929, which could effectively improve the objectivity and accuracy of image quality assessment [16]. Low-resolution images often fail to provide sufficient data for analysis when containing scene-related information such as lighting, shadows, and reflections, thereby limiting the ability to extract effective information from the image. The challenges faced by low image resolution are particularly evident in image processing and reconstruction. Improving image quality through intelligent technology and next-generation super-resolution algorithms remains a key research focus, especially in addressing dynamic environments and hardware limitations.

At the same time, CNN image research is relatively mature. Zhang et al. proposed an infrared and visible image fusion algorithm based on ResNet-152 to extract multi-layer features by decomposing the low-frequency and high-frequency parts of the image. Experiments showed that the proposed method was superior to the comparison algorithm in retaining important features and obtaining more details [17]. Shao G et al. proposed a subpixel CNN for image super-resolution reconstruction, converting RGB to YCbCr mode, and introducing residual networks and upsampled subpixel convolutional layers (CLs) for improvement. Experimental results showed that it outperformed many traditional methods in terms of

accuracy and time consumption [18]. Jin Z proposed a flexible deep CNN framework for processing in computer vision, which utilized the frequency characteristics of different types of workpieces. By adjusting the architecture, the same method could be utilized for various IR tasks. The results illustrated that the framework was more excellent than other methods [19]. Zhang Y designed a universal image fusion model for CNN. This study first utilized two CLs to extract image features from the II. Then, the features of the II were fused by selecting appropriate fusion rules. The results showed that compared with other fusion models, the research model had better pan-Chinese ability and achieved better fusion results [20]. Zhu F et al. proposed a CNN-based denoising method for better obtaining clean images from noise. Combining some different rate expansion convolutions with common convolutions enriched the features extracted from multi-layer convolutions. The results showed that the denoising effect of the research method was excellent [21]. To improve the edge detection performance of noisy images, Yuan S et al. proposed an edge detection method based on nonlinear structural tensors, which determined image edges by calculating the tensor product, eigenvalues, and eigenvectors of noisy images. The results indicated that this method could detect more monitoring points compared to other methods, with a shorter average detection time and overall better detection performance [22].

In summary, DM has brought convenience to people's lives. There is relatively abundant research on visual Image Design (ID) and image reconstruction, but there is relatively little research on CNNs. The development of CNN is relatively mature and has achieved good results in image denoising, feature extraction, and detection. In view of this, this study proposes a new CNN-based DM visual ID and image reconstruction method.

3 ID and image reconstruction method for CNN algorithm

To improve the reconstruction quality of DM visual images, a SRCNN dual-channel reconstruction algorithm is proposed, which adds guided channels and uses masked images for improvement. The improvement and innovation of this study is the utilization of explicit edge mask-guided parallel channels to provide powerful structural priors for super-resolution reconstruction. Its mechanism is essentially different from traditional implicit feature learning or simple multi-scale fusion, and can greatly improve the quality of image reconstruction.

3.1 SRCNN reconstruction algorithm

There are various methods for image super-resolution, including instance-based, interpolation-based, and reconstruction-based methods and deep learning algorithms [23]. In CNN algorithms, SRCNN is a representative algorithm applied to image super-resolution, which can learn the unique connections between high-resolution and low-resolution images and use these connections as constraints to generate high-resolution images. Research has shown that using multi-layer network structures can enable models to learn more complex features and patterns, which is particularly important for generating high-quality images and reconstruction tasks [24]. Figure 1 shows the relevant framework.

As shown in Figure 1, the three CLs are utilized for feature extraction, nonlinear mapping, and high-resolution image reconstruction. The entire process is to first extract small blocks from the interpolated low-resolution image and represent the extracted small blocks in the form of high-dimensional vectors. Then, it transforms all high-dimensional vectors nonlinearly and maps them onto other high-dimensional vectors [25]. Finally, it aggregates all high-resolution small blocks to form the final high-resolution image, which should be similar to the ideal high-resolution image corresponding to the low-resolution image. The SRCNN algorithm mainly uses convolution operations to train and process input images during forward and backward propagation. The CL can convert digital signals into low-dimensional vectors and store the extracted features in the feature vectors. The convergence rate of the algorithm is restricted by the selection of the Loss function. This study uses the fast convergence cross-entropy as the loss function and adds a minimum value to it, as shown in equation (1).

$$\begin{cases} \hat{y}' = \hat{y} + \sigma \\ C = -\left[y \ln \hat{y}' + (1-y)\ln(1-\hat{y}') \right] \end{cases} \tag{1}$$

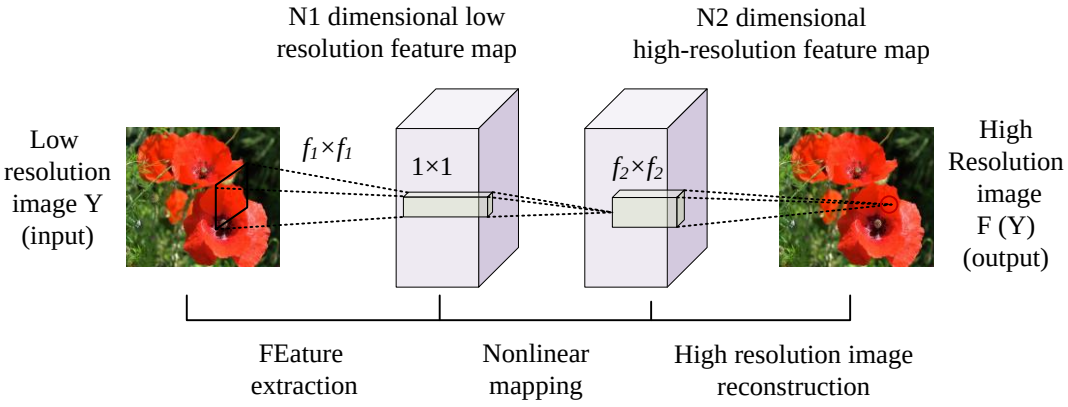


Figure 1: The network framework of SRCNN.

As shown in equation (1), the predicted output is represented by $\hat{y}$. The actual output is represented by y. σ represents the minimum value. The commonly used CNN algorithm is aimed at the gradient descent method. Although it has good performance, it also has disadvantages, such as difficulty in choosing an appropriate learning rate, and easy occurrence of saddle points and local extremum problems [26]. Therefore, an adaptive learning rate is selected to overcome the shortcomings. The Adam algorithm is selected in this study. The Adam algorithm can dynamically adjust the learning rate of each parameter, and it can retain historical gradient information, accelerate the convergence process, and reduce oscillation phenomena. The random gradient descent algorithm for the Adam algorithm is shown in equation (2).

$$\begin{cases} m_t = \rho_1 * m_{t-1} + (1-\rho_1) * g_t \\ n_t = \rho_2 * n_{t-1} + (1-\rho_2) * g_t^{\,2} \\ \hat{m}_t = \dfrac{m_t}{1-\rho_1^t} \\ \hat{n}_t = \dfrac{n_t}{1-\rho_1^t} \\ \Delta\theta = -\dfrac{\hat{m}_t}{\sqrt{\hat{n}_t} + \delta} * \eta \\ \theta_{t+1} = \theta_t + \Delta\theta \end{cases} \tag{2}$$

In equation (2), the gradient is represented by g_t. ρ_1 and ρ_2 are the decay rates. m is the first-order moment of the gradient. $\hat{m}_t$ is the moment estimate after deviation correction of the first-order moment of the gradient. n_t is the second-order moment of the gradient. $\hat{n}_t$ is the moment estimate after deviation correction of the second-order moment of the gradient. Learning rate is expressed in η. δ represents the minimum constant. A large number of pooling operations may result in blurry feature information. Therefore, a *soft* max function classifier is added after the fully connected layer to convert complex result values into relative probabilities that are easy to compare and understand. There are a large number of shared layers between different neurons, which can effectively reduce the amount of data. Research has been conducted to improve traditional neurons by using random deactivation, as shown in equation (3).

$$\begin{cases} r^{(l)} \sim Bernoulli(p) \\ \tilde{a}^{(l)} = r^{(l)} * a^{(l)} \\ c^{(l+1)} = w^{(l+1)} \tilde{a}^{l} + b^{(l+1)} \\ a^{(l+1)} = f(c^{l+1}) \end{cases} \tag{3}$$

As shown in equation (3), the selected neuron is represented by $\tilde{a}^{(l)}$; The sample subset of Bernoulli distribution satisfying probability p is represented by $r^{(l)}$. $a^{(l+1)}$ represents $l+1$ layer output. $l+1$ input weighting and vector are represented by $c^{(l+1)}$. The weight of $l+1$-layer is represented by $w^{(l+1)}$. $b^{(l+1)}$ is the $l+1$-layer bias. $f()$ is the excitation function.

3.2 Design of Improved SRCNN Dual-Channel Mask-Guided Network Architecture Based on Image Reconstruction

3.2.1 Overall Network Architecture

Although SRCNN has good recognition and image reconstruction performance, its network structure relies heavily on its own neural network structure, resulting in limited learning ability. Therefore, to improve SRCNN, this study introduces a parallel guidance channel and utilizes edge masks explicitly extracted from the input image as structural priors to guide the feature learning process of the main channel. The purpose is to enhance the network's ability to recover high-frequency details and further improve the reconstruction performance of the neural network. The improved algorithm structure is shown in Figure 2.

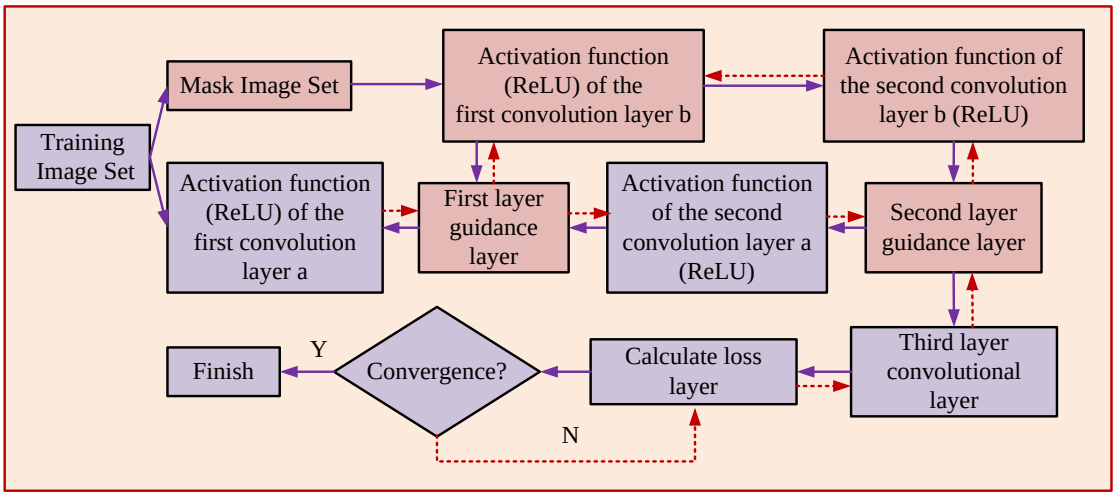


Figure 2: Improved SRCNN algorithm structure.

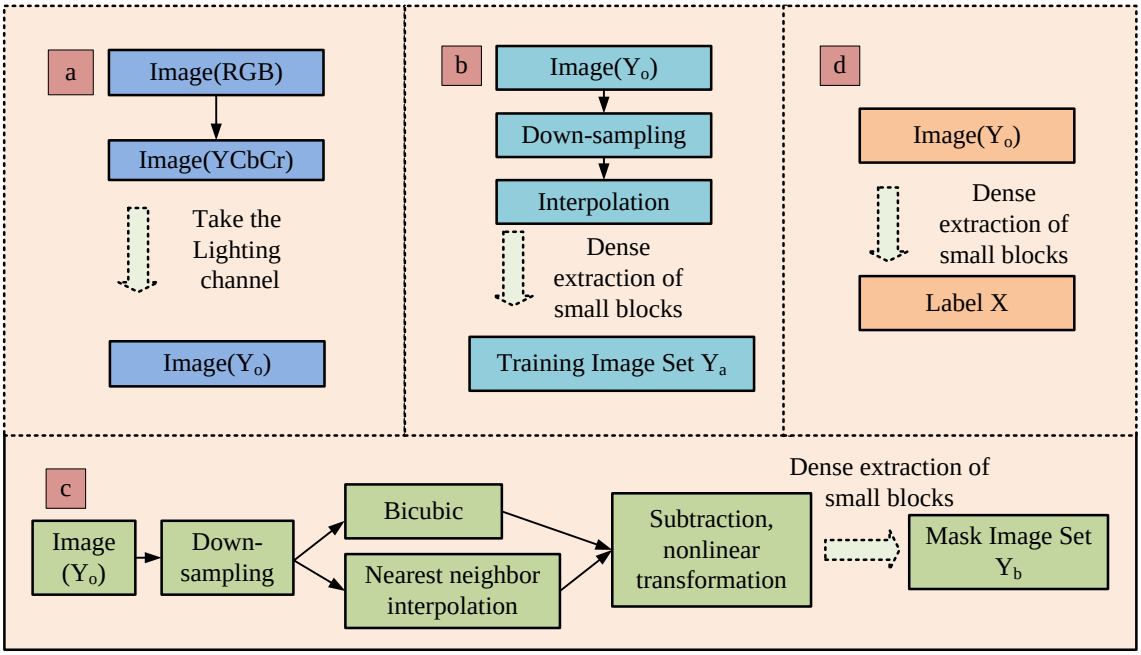


Figure 3: The process of obtaining a training image set.

As shown in Figure 2, the improved network has three CLs and two channels. The solid line represents forward conduction, while the red dashed line represents reverse conduction. The red part represents the improved part, which is the newly added second channel. After adding channels, five modules are added, namely the first guiding layer, the second guiding layer, the first CL, the second CL, and the mask image dataset. By introducing additional guidance channels and using masked images of input images, new dimensions can be added to feature learning, thereby improving network efficiency. Specifically, the guidance channel can provide additional information at different scales related to the main input image, thereby enhancing the network's perception ability of different features. Mask images are used to emphasize important features such as edges, textures, and structures in the image, and utilize these features to guide the original convolutional channels.

3.2.2 Mask generation module

To provide clear edge guidance information to the network, a corresponding mask image is first generated for each input image. The specific method is to use different interpolation algorithms to address the differences in edge processing characteristics. Firstly, the original low-resolution image is interpolated using bicubic interpolation and nearest neighbor interpolation to obtain two images, which are then subtracted and subjected to a nonlinear transformation to enhance the edges. The image training method can be specifically shown in Figure 3.

In Figure 3, firstly, the images obtained from the training set are transferred from the RGB color space to the YCbCr color space, and the image Y0 is obtained through the illumination channel. During image downsampling, after double triple interpolation, overlapping small blocks are extracted from the interpolated image, and then the extracted small blocks are aggregated to form a training image set Y_a. Bicubic interpolation and nearest neighbor interpolation are used to upsample the original low-resolution image, and then the two resulting images are subtracted and the edges of the image are enhanced through nonlinear transformation. This process can be represented by equation (4).

$$Y_{mask} = \tanh(\beta * | f_b(Y_{ds}) - f_n(Y_{ds}) |) \quad (4)$$

As shown in equation (4), the mask image is represented by Y_{mask}. The sampled image is represented by Y_{ds}. Bicubic interpolation is represented by f_b. The nearest neighbor interpolation is represented by f_n. The transformation size coefficient of nonlinear transformation is represented by β, which is generally negative. tanh() is the hyperbolic tangent activation function, which normalizes the mask values between (-1,1) to make the edge features more prominent. By performing differential operations on bicubic interpolation (preserving smooth regions) and nearest neighbor interpolation (highlighting edge step effects), high-frequency edge information in the image can be enhanced, mask image M can be generated, and the network can be guided to focus on structural features. Traditional SRCNN relies solely on CL learning for low-resolution to high-resolution mapping, lacking explicit utilization of inherent image features such as edges and textures. The addition of a guiding channel can optimize the input mask image through dual-channel collaboration, preserving its details and improving its image reconstruction effect. The acquisition of labels is shown in Figure 4.

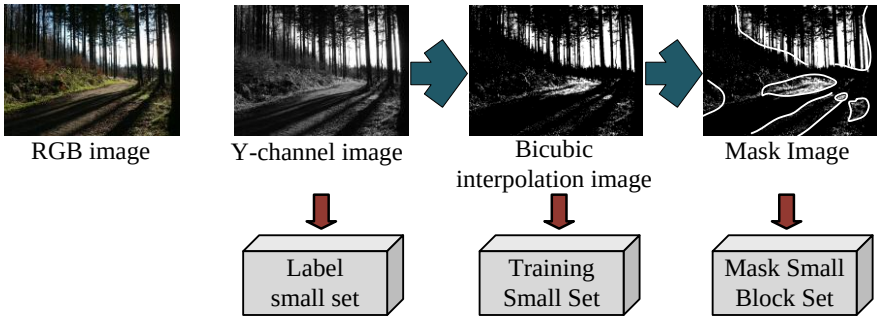


Figure 4: Acquisition of labels.

Figure 4 shows the specific extraction process of label X. This is a randomly selected image from the training set, displaying mask image blocks, training image blocks, and label blocks, respectively. The purpose of the first CL is feature extraction, extracting small blocks from images Y_a and Y_b to obtain II feature $F_1(Y)$, which is similar to the case through a set of filters, and its expression is shown in equation (5).

$$\begin{cases} F_{1a}(Y) = \max(0, W_{1a} * Y_a + B_{1a}) \\ F_{1b}(Y) = \max(0, W_{1b} * Y_b + B_{1b}) \end{cases} \quad (5)$$

As shown in equation (5), the two channels are denoted by a and b. $F_{1a}(Y)$ is the content feature map output by the main channel in the first layer, and $F_{1b}(Y)$ is the structural feature map output by the guiding channel in the first layer. The convolution kernels (CKs) of the two channels are represented by W_{1a} and W_{1b}, respectively. Y_a and Y_b represent training images and mask images, respectively. $*$ represents convolution operation. The bias vectors of the two channels are represented by B_{1a} and B_{1b}

3.2.3 Dual-channel feature extraction and fusion

After the mask is generated, the upsampled image and the mask image are fed into the main and guide channels, respectively. The main channel is responsible for learning the overall content and texture information of the image, and the guide channel is responsible for learning the structure and edge information of the image. The dual-channel separation design can avoid feature confusion and provide prior structural constraints. The purpose of the second layer of CL is nonlinear mapping, that is, to transform the feature vector of low-resolution space into high-resolution space. At this time, the output feature is $F_2(Y)$, and its expression is shown in equation (6).

$$\begin{cases} F_{2a}(Y) = \max(0, W_{2a} * G_1(Y) + B_{2a}) \\ F_{2b}(Y) = \max(0, W_{2b} * F_{1b}(Y) + B_{2b}) \end{cases} \quad (6)$$

As shown in equation (6), the CK of two channels are represented by W_{2a} and W_{2b}, respectively. The output of the first guidance layer is represented by $G_1(Y)$. The output of the first CL is represented by $F_{1b}(Y)$. The bias vectors of the two channels are represented by B_{2a} and B_{2b}. The nonlinear mapping of the second layer convolution can map low-dimensional features to high-dimensional space and fuse dual-channel information. Guide channel features can enhance edge consistency and prevent reconstruction blurring. The purpose of the third layer convolution is to reconstruct high-resolution image. The final generated high-resolution image is $F(Y)$, and its expression is shown in equation (7).

$$F(Y) = W_3 * G_2(Y) + B_3 \quad (7)$$

As shown in equation (7), the CK is represented by W_3. The output of the second guidance layer is represented by $G_2(Y)$. The bias vector is represented by B_3. The third layer does not need to undergo nonlinear transformation, as the W_3 of the third layer can be regarded as a filter, and the reconstruction process can be considered as a linear transformation process. The mask image in Figure 4 shows that it can represent edge features, and these features are used to guide the convolutional channel. The formula for the guidance layer is shown in equation (8).

$$G_i(Y) = F_{ia}(Y) \times F_{ib}(Y) \quad i = 1, 2 \quad (8)$$

As shown in equation (8), $F_i(Y)$ represents the output image of the i-th CL. The research method requires training parameters $\{W_{1a}, W_{1b}, W_{2a}, W_{2b}, W_3, B_{1b}, B_{2b}, B_3\}$, and the training process is actually a part of parameter optimization and estimation. By reducing the loss between high-resolution image and reconstructed images, the optimal solution of the parameters is found. The visualization of guide layer features validates the effective utilization of edge information. To train the network parameters, the study aims to optimize the model by minimizing the mean square error (MSE) between the reconstructed image and the real high-resolution image, and its expression is shown in equation (9).

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \left\| F(Y_i; \Theta X_i) \right\|^2 \quad (9)$$

As shown in equation (10), the reconstructed image is represented by $F(Y_i; \Theta X_i)$. The number of training samples is represented by n. The high-resolution image set is represented by $\{X_i\}$. The set of processed images with the same size as the original image is represented by $\{Y_i\}$.

4 Result analysis of ID and image reconstruction methods for CNN algorithm

4.1 Experimental design platform and related parameter settings

To evaluate the proposed CNN-based DM visual ID and image reconstruction method, multiple IG tasks are selected in this experiment. It includes facial recognition, facial IR, IG, and IR, with multiple publicly available datasets. This study first selects the parameters of the CNN and then tests and verifies the performance of the algorithm. To avoid experimental errors caused by different equipment, the study uses the same computer for the experiment, using Intel Core i7-9700 3.00GHz CPU and 20GB memory. Operating System: Ubuntu 20.04 LTS, with the deep learning framework of PyTorch 1.10. The core library is CUDA 11.3, Python 3.8, and OpenCV 4.5. The experiment uniformly adjusts the input image to 256 × 256 pixels (based on common preprocessing standards such as CIFAR-10 and ILSVRC2020) to meet the requirements of multi-scale feature extraction. The size of the first CK is 9×9 (large kernel captures global structure), the size of the second CK is 6×6 (moderate kernel balances details and computational complexity), and the size of the third CK is 5×5 (small kernel refines local features). Stride defaults to 1 to ensure high-resolution feature preservation (applicable to all CLs). In addition, in the second layer of the guidance channel, Stride=2 is used to downsample the mask image to reduce computational complexity. The AF of the first- and second-layer channels is LeakyReLU (α=0.2), which alleviates the problem of gradient vanishing. The third layer channel is linearly activated (without AF), as the reconstruction task requires retaining all intensity values. ReLU is used throughout the guidance channel to enhance the sparsity of edge features. The initial learning rate is 0.0001, and the optimizer is Adam (β_1=0.9, β_2=0.999, and ε=1e-8). Image blocks of 96*96 pixels are randomly cropped from the high-resolution image. The corresponding low-resolution image block is obtained by performing bicubic downsampling (magnification factor x3) on the high-resolution image block. To increase data diversity and improve model generalization ability, online data augmentation is applied to the training data, including random horizontal flipping and 90-degree rotation. The optimizer uses improved Adam, and the initial learning rate is set to 1e-4. After every 20 rounds, the learning rate decays to 0.5 times the original value. Table 1 shows the experimental parameter settings.

4.2 Dataset source and related information

The CIFAR-10 dataset and ILSVRC2020 dataset have a wide range of applications and can be used to evaluate image processing and reconstruction algorithms. The CIFAR-10 dataset contains a variety of color image categories and has rich visual effects, which can effectively test the algorithm's processing effect on visual images. The ILSVRC2020 dataset has a large amount of data, rich sample size, and high image resolution, making it more suitable for complex scenes and high detail testing content. The use of CIFAR-10 and ILSVRC2020 datasets to evaluate the performance of ID and image reconstruction algorithms is due to their widespread application and recognized benchmark position in the field of computer vision. Moreover, their diversity, rich content, differences in data volume and structure, as well as balanced data distribution, facilitate the comprehensive evaluation of algorithm performance and robustness from multiple aspects. This study selects 100 images from the CIFAR-10 and ILSVRC2020 datasets for testing experiments. 100 images are randomly selected from CIFAR-10 and ILSVRC2020 as the test set, and the remaining data are used for training. The extraction of the test set adopts stratified sampling, where CIFAR-10 is uniformly selected into 10 categories, with 10 images per category. ILSVRC2020 is grouped according to the semantic hierarchy of ImageNet (such as animals and artificial objects), with 20 images extracted from each group. All experiments are repeated 5 times, with random seeds fixed at {42, 123, 2023, 55, 17}, and the results are taken as mean ± standard deviation.

4.3 Parameter experiment of CNN network structure

For selecting an appropriate size of CK, the interpolation ratio is 3, the learning rate is 0.0001, and the quantity of CK in the first layer is 64. The quantity of CK in the second layer is 32, and an unimproved CNN algorithm is introduced as a reference to compare the loss values (LVs) of reconstructed images with CK sizes of 9-2-5, 9-4-5, and 9-6-5, respectively. The results are shown in Figure 5.

Table 1: Experimental parameter settings.

Parameter	Numerical value
Batch Size	16
Epochs	100
Loss function	MSE + 0.5×MSSIM
Weight initialization	He follows a normal distribution
Decline in learning rate	2
CK size	9-6-5
Guide channel downsampling Stride	2 (second layer only)

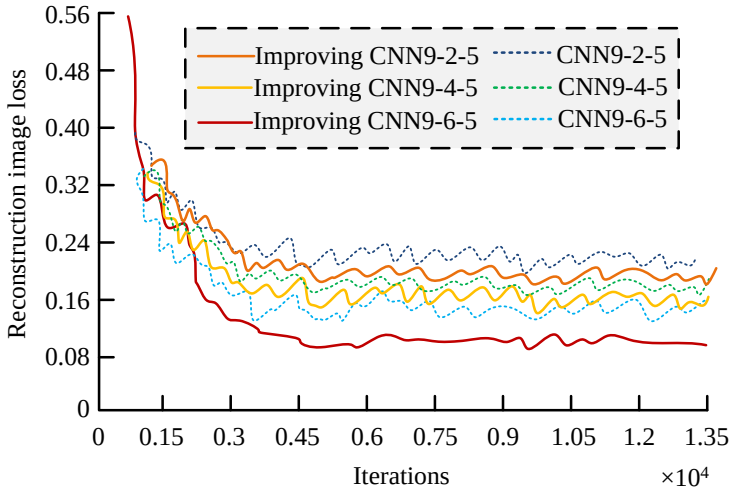


Figure 5: Reconstruction image LV for different CK sizes.

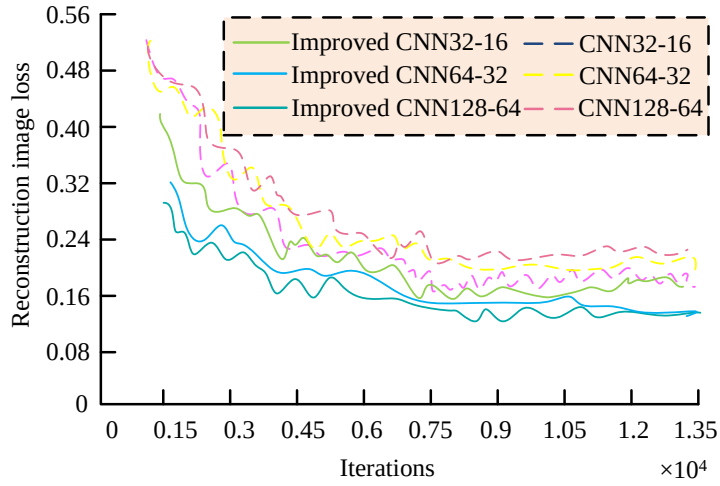


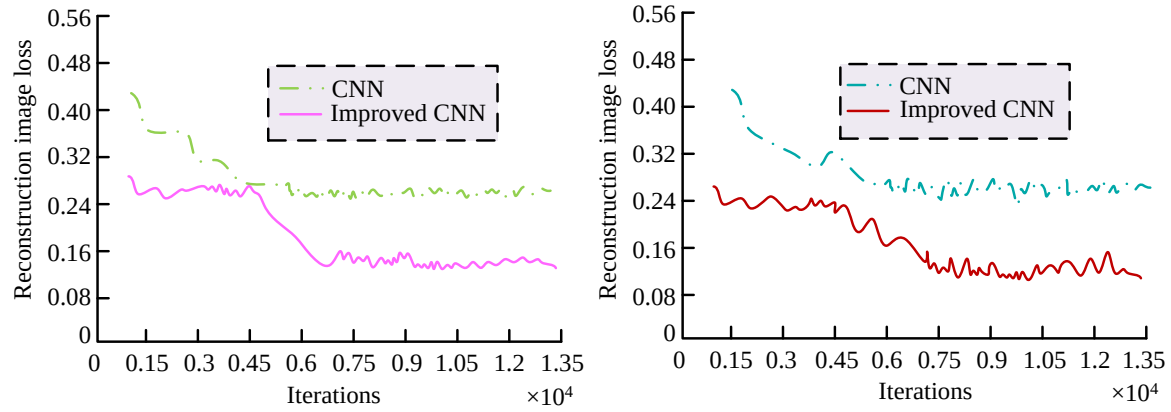
Figure 6: Loss of reconstructed images with different CK numbers.

Figure 5 shows the loss result graph plotted before 13,500 iterations. Compared with the unmodified CNN, the improved CNN has a significantly faster rate of convergence. At the iteration number of 4,000, the CNN of each size has a significant convergence, and the LVs are 0.10, 0.16, and 0.20 respectively, while the unmodified CNN of each size has an LV of 0.15, 0.18, and 0.24, and does not converge to the minimum value. Figure 5 also indicates that as the size of the second CK increases, the LV of the reconstructed image gradually decreases. This indicates that an increase in the size of the second CK will enable the neural network structure to learn more information, reduce LV, and achieve better ID and image reconstruction results. The number of CK also affects the experimental results. Choosing the same parameters as above, the fixed CK size is 9-6-5, the quantity of CK is 128 in the first layer, 64 in the second layer, 64 in the first layer, 32 in the second layer, 32 in the first layer, and 16 in the second layer. The LV of reconstructed images with different CK numbers is shown in Figure 6.

As shown in Figure 6, as the number of CK increases, the neural network structure learns more information, the LV of the reconstructed image is lower, and the image effect is better. Compared to the unmodified CNN, the improved CNN converges earlier. When the quantity of convolutions in the first layer is 128, the quantity of convolutions in the second layer is 64, and the quantity of iterations is 7,000, the improved CNN convergence LV is 0.15. At this point, the LV of the unimproved CNN is 0.23 and has not yet converged to the minimum value.

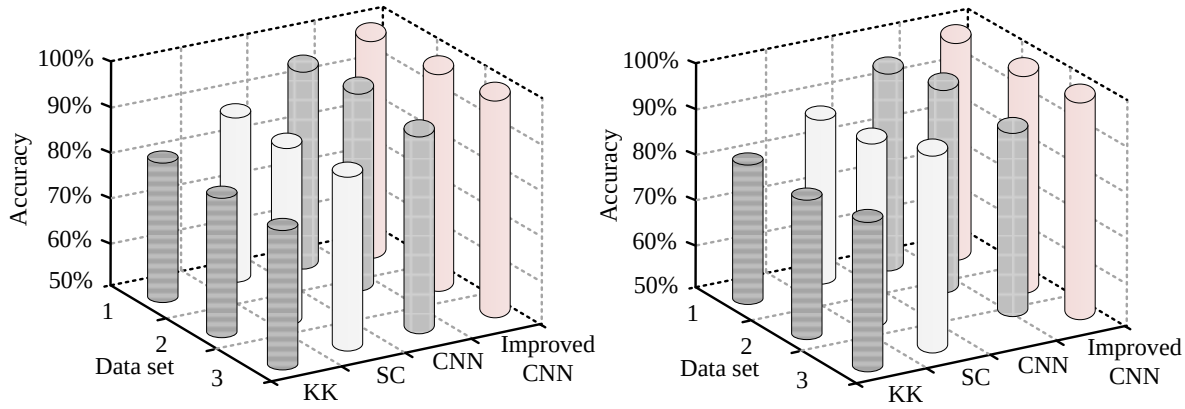
4.4 Comparative analysis experiment of ID and image reconstruction methods for CNN

To verify the superiority of DM visual ID and image reconstruction methods over CNN, an unimproved CNN algorithm is further introduced for comparison. At this point, the CK size is set to 9-6-5 and the quantity of CK in the first layer is 128, while the quantity of CK in the second layer is 64. The comparison results are shown in Figure 7.



(a) CIFAR-10 (b) ILSVRC2020

Figure 7: Comparative experiments of two datasets.



(a) CIFAR-10 (b) ILSVRC2020

Figure 8: Comparison of accuracy of four methods.

Figure 7 (a) shows that in the dataset CIFAR-10, CNN converges at 4,500, leading to an LV of 0.25. The improved CNN converges at 7,500, leading to an LV of 0.16. Figure 7 (b) shows that in the dataset ILSVRC2020, CNN converges at 6,000, leading to an LV of 0.29. The improved CNN converges at 7,000, leading to an LV of 0.10. This indicates that the improved research method results in lower LV and better results in obtaining the final design and reconstruction images. To further validate the superiority of the research algorithm, image reconstruction methods are introduced: sparse regression and natural image prior (KK) and sparse coding-based method (SC) are used to compare the accuracy, error, response time, and image evaluation results of the four methods. In the CIFAR-10 dataset, images are divided into high, medium, and low layers based on their edge complexity (gradient amplitude calculated using the Sobel operator). 30 images are randomly selected from each layer to a ensure balanced distribution of structural complexity among the three groups. The accuracy of the four methods is shown in Figure 8.

Figure 8 shows that the improved CNN is superior to the other three algorithms in accuracy on both datasets. The accuracy of the algorithm is ranked from high to low as Improved CNN, CNN, SC, and KK. In Figure 8 (a), the improved CNN method achieves the highest accuracy of 97.23% and an average of 95.23% for the four sets of data. In Figure 8 (b), the improved CNN method achieves the highest accuracy of 98.23% and an average of 96.25% for the four datasets. Overall, the accuracy of the improved CNN performs well on both datasets. The mean absolute percentage error (MAPE) of the four methods is shown in Figure 9.

In Figure 9(a), the MAPE of the 20 images of the SC and KK algorithms is 3.92% and 4.12%. The improved CNN curve fluctuated the most, with a MAPE of 1.02% for 20 images. In Figures 9(a) and (b), the improved CNN has minimal MAPE. The proposed method includes guide channel and mask images, which can effectively extract image edge features and enhance feature extraction capabilities. However, traditional sparse coding or regression methods are often based on fixed prior information and lack flexibility. Sparse regression and natural image priors often rely on linear combinations, which makes it difficult to capture complex features. The improved CNN has good accuracy, error rate and response time. Ten photos are selected from the two datasets for testing. The response times of the four methods are shown in Figure 10.

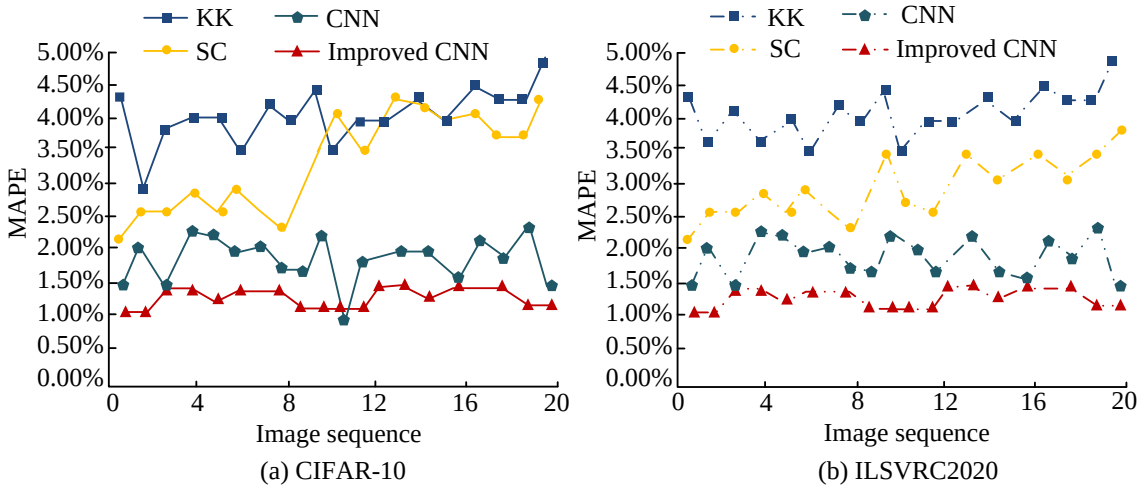


Figure 9: Comparison of MAPE of four models.

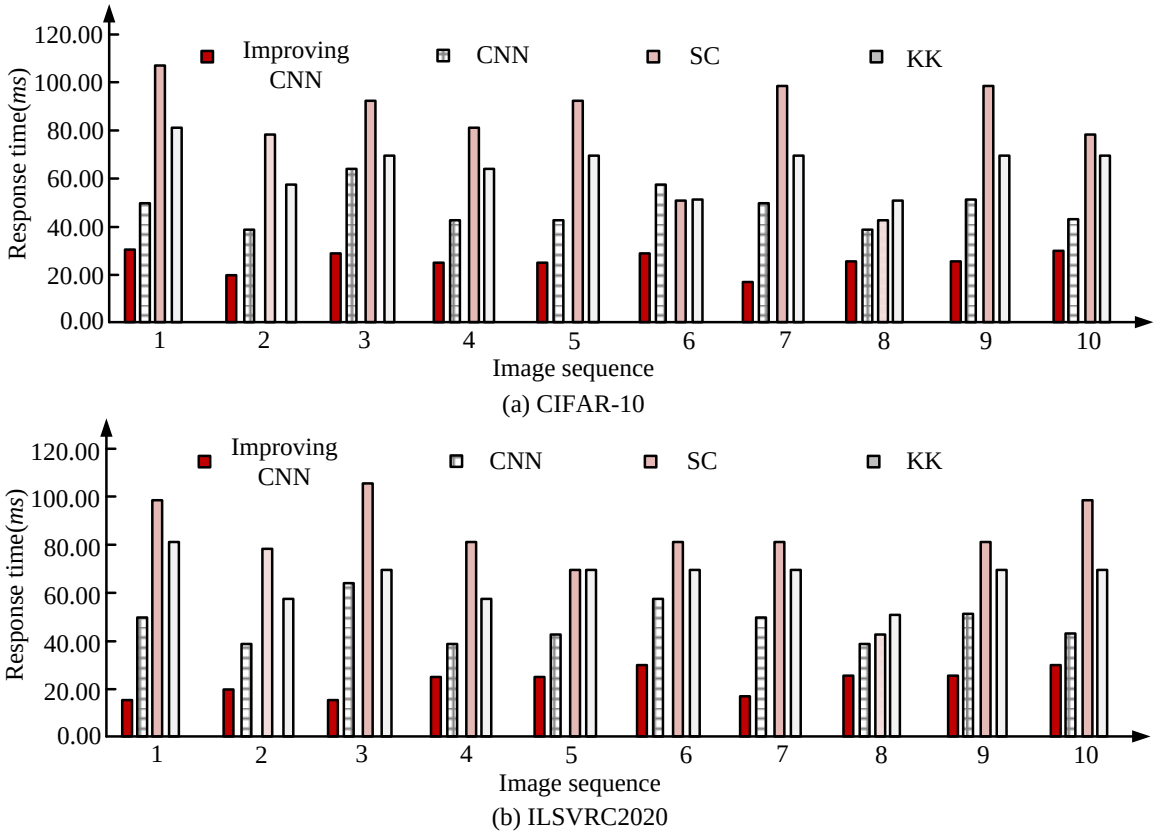


Figure 10: Comparison of response times of four methods.

Figure 10 shows that the response time of the improved CNN is lower than that of the other three algorithms. The response times of these algorithms from low to high are improved CNN, CNN, KK, and SC. Figure 10 (a) shows that the minimum response time of the improved CNN method is 18 *ms* and the average value is 21 *ms*. Figure 10 (b) shows that the minimum response time of the improved CNN method is 16 *ms*, and the average value is 20 *ms*. The improvement research scheme can coordinate the learning process of the guide channel and the main input channel more effectively by jointly optimizing the loss function of the guide, thereby significantly improving the reconstruction effect. The simple integration of the boot channel does not significantly increase the computational complexity, and its precise grasp of key information significantly enhances performance while maintaining responsiveness. To further analyze the computational efficiency of the research method, the performance differences between the research method and CNN, KK, and SC methods are compared in terms of parameter size, inference time, FLOPs, and other indicators. The results are shown in Table 2.

Table 2: Comparison of the computational efficiency of different algorithms.

Index	Parameter (M)	FLOPs (G)	GPU Memory (GB)	Inference time (ms)	PSNR (dB)
KK	0.27	0.8	0.5	35	28.50
SC	0.12	2.3	1.2	45	28.21
CNN	0.24	4.7	1.8	25	30.05
Improved CNN	0.38	3.5	1.6	20	32.52

Table 3: Comparison of image quality evaluation results of four methods.

Evaluation criterion	Interpolation Scale	Double triple interpolation	KK	SC	CNN	Improved CNN
MSSIM	2	0.9976	0.9932	-	0.9975	0.9981
	3	0.9738	0.9769	0.9801	0.9814	0.9875
	4	0.9517	0.9687	-	0.9712	0.9789
WPSNR	2	50.12	57.02	-	58.97	60.64
	3	41.19	46.87	47.12	47.19	47.99
	4	37.33	40.48	-	47.95	41.98
NQM	2	36.97	38.9	-	41.23	43.98
	3	27.75	32.87	33.04	33.17	34.54
	4	21.86	24.34	-	25.65	26.87
IFC	2	6.19	6.46	-	8.21	8.47
	3	3.53	3.89	4.57	4.79	5.01
	4	2.23	2.56	-	3.74	3.88
SSIM	2	0.9268	0.9521	-	0.9578	0.9682
	3	0.8648	0.8711	0.8864	0.9068	0.9369
	4	0.8148	0.8518	-	0.8647	0.8738
PSNR	2	33.21	36.89	-	36.98	37.12
	3	30.23	31.51	32.87	33.86	34.97
	4	28.75	30.48	-	30.89	31.97

In Table 2, the number of parameters (0.38M) of the proposed improved CNN is increased by 58% compared with the original CNN, which is mainly due to the CK of the new guide channel. The improved CNN can control the growth through the parameter sharing strategy (such as the mask channel shares some weights with the main channel). The FLOPs of the improved CNN are reduced by 25% (3.5G vs 4.7G), which benefits from the sparsity of mask guidance and skips the low-contribution region in the second layer convolution. The inference speed (20 ms) of the improved CNN is better than that of SC (45 ms) and CNN (25 ms), and the memory occupation (1.6GB) is between SRCNN (1.8GB) and SC (1.2GB). The improved CNN has better reconstruction quality (PSNR: 32.52 dB), which is 2.47 dB higher than that of the original SRCNN. KK is parameter-dependent, but has a low PSNR (28.50 dB) because it cannot learn data-driven features. SC needs iterative optimization, the inference speed is slow (45ms), and it is sensitive to occlusion. The improved CNN can achieve quality-efficiency balance under the condition that the number of parameters is limited. To further validate the applicability of the research method, image quality evaluation indicators such as PSNR, SSIM, information fidelity criterion (IFC), noise quality measure (NQM), weighted PSNR (WPSNR) are introduced. The multi-scale SSIM (MSSIM) evaluates image quality. The IFC considers the original image as the source of information and the distortion process as an information channel. It evaluates the image quality and the fidelity of information transmission by calculating how much information can be extracted from the reconstructed image. IFC has a strong correlation with human perception of the naturalness and clarity of images. NQM works by simulating the sensitivity of the human visual system to different spatial frequencies, effectively quantifying annoying distortions such as block effects and ringing effects introduced by compression or reconstruction algorithms. The lower the value of NQM, the less perceived noise and higher quality of the image. WPSNR assigns different weights to different regions when calculating errors, and better reflects the reconstruction quality of visually important regions than standard PSNR. Under the optimal parameter settings, 20 photos are randomly selected from the CIFAR-10 data for application, and the outcomes are indicated in Table 3. The 20 samples for image quality evaluation include 5 high texture, 5 low texture, 5 high dynamic range, and 5 occluded scenes.

In Table 3, the MSSIM value of the improved CNN is greater than 0.97 under different interpolation ratios, while the MSSIM value of the KK and CNN is slightly worse. The improved CNN achieved WPSNR values of 60.64, 60.64, and 41.98 at interpolation ratios of 2, 3, and 4, respectively, which was still higher than other comparison algorithms under the same conditions. The NQM values (43.98, 34.54, 26.87) of the improved CNN outperforms the other methods in all cases, and the IFC value is stable. The above results indicate that the proposed algorithm achieves higher similarity in image content reconstruction results at different scales, and performs better than other algorithms in feature extraction and image reconstruction. Among them, the SC performs the worst, as it is difficult to grasp image details and structural content solely through sparse encoding methods. The KK performs better than the SC, but its prior thinking based on natural images is difficult to ensure structural similarity and reconstruction results. To further evaluate the performance of the improved CNN based on dual-channel guidance, it is compared with the residual channel attention network (RCAN), enhanced deep super

resolution (EDSR), super-resolution generative adversarial network (SRGAN), dense residual network (DRN), and Swin Transformer for Image restoration (SwinIR) on the CIFAR-10 dataset. The interpolation ratio is 3, and 50 test images are randomly selected for quantitative evaluation. The comparison results are shown in Table 4.

In Table 4, SwinIR performs the best on most objective metrics (PSNR, SSIM, IFC, NQM, and WPSNR), thanks to the powerful global modeling capability of the Transformer architecture. The PSNR and SSIM of the improved CNN are only 0.15 dB and 0.0016 lower than SwinIR, proving its good reconstruction quality. Improved CNN outperforms all comparison methods in MSSIM metrics (0.9875), indicating its advantage in maintaining multi-scale structural similarity. Its parameter size is only 1.2M, far lower than complex networks such as EDSR (43.1M) and DRN (22.3M), demonstrating good parameter efficiency. Improved CNN's 21.3 ms inference time is about 45% faster than EDSR and about 66% faster than SwinIR, resulting in a significant reduction in computational cost. RCAN improves performance by channel attention, but has a large number of parameters (156 m). To further validate the improved CNN, it is compared with the Channel Spatial Hybrid Attention Network for Image Super Resolution (CSHA), Multi-level Information Compensated U-Net for Image Super Resolution (MICHNet), SRGAN, and Soft Edge-guided Progressive Super Resolution Network (SEPNet) in terms of computational efficiency, reconstruction quality, resource consumption, and reconstruction accuracy. The comparison results are shown in Table 5.

In the results of Table 5, the research method maximizes the preservation of the structural information of the original image due to the dual-channel design (main channel+mask guidance), resulting in the best information fidelity (8.47) compared to CSHA (7.82), MICHNet (7.35), SRGAN (6.21), and SEPNet (8.12). Although SEPNet adopts edge compensation strategy (IFC=8.12), it does not explicitly model the inter channel dependencies, which is slightly inferior to the research method. The NQM value of the research method (43.98) performs well, increasing by 9.6% compared to CSHA (40.12), while SRGAN has the lowest NQM value (35.44) due to the introduction of high-frequency artifacts in adversarial training. The WPSNR value of the research method (60.64 dB) verifies the effectiveness of mask guidance, while MICHNet (56.91 dB) performs poorly in areas with complex textures due to feature loss issues. The MAPE of the research method is the lowest (1.00%), a decrease of 10.7% compared to SEPNet (1.12%), while the MAPE of SRGAN reaches 2.32%, mainly due to systematic biases in the generated images (such as color shift). The reconstruction quality of the research method performs well, with an SSIM value (32.52 dB), 0.72 dB higher than CSHA and significantly better than SRGAN (+2.32 dB). The reason for this is that the improved CNN explicitly enhances edge features through mask-guided channels, while CSHA only indirectly allocates resources through attention weights. SRGAN is prone to artifacts due to GAN. The above structure indicates that the improved CNN is comprehensively leading in perception indicators such as IFC, NQM, and WPSNR. This indicates that the improved CNN can improve pixel accuracy while also better maintaining visual naturalness and balancing computational efficiency to a certain extent. Afterwards, the image reconstruction effects of the above comparison methods are presented, and the results are shown in Figure 11.

The results in Figure 11 indicate that the image clarity trained by the improved CNN is significantly higher than other comparison models, with higher image clarity and better restoration of image details. Secondly, the SEPNet and CSHA perform well. Afterwards, the comparison results of different algorithms are analyzed using the statistical tests in Table 6.

Table 4: Image quality evaluation results of different comparison methods.

Evaluation metric	RCAN (ECCV'18)	EDSR (CVPR'17)	SRGAN (CVPR'17)	DRN (CVPR'18)	SwinIR (ICCV'21)	Improved CNN
PSNR (dB)	34.21	33.98	32.45	34.35	35.12	34.97
SSIM	0.9321	0.9287	0.9123	0.9338	0.9385	0.9369
IFC	4.89	4.76	4.21	4.95	5.12	5.01
NQM	33.87	33.45	31.89	34.02	34.78	34.54
WPSNR (dB)	47.45	47.12	45.89	47.56	48.21	47.99
MSSSIM	0.9845	0.9832	0.9789	0.9851	0.9867	0.9875
Parameter quantity (M)	15.6	43.1	1.5	22.3	11.9	1.2
Inference time (ms)	45.2	38.7	52.3	41.5	62.8	21.3
Training time (m)	48	36	72	54	96	24

Table 5: Comparison of reconstruction performance of different algorithms.

Model	PSNR (dB)	SSIM	IFC	NQM	WPSNR (dB)	MAPE (%)	FLOPs (G)	Inference time (ms)	GPU Memory (GB)
Improved CNN	32.52	0.9682	8.47	43.98	60.64	1.00	3.5	20	1.6
CSHA	31.80	0.9621	7.82	40.12	58.23	1.45	8.7	35	2.8
MICUNet	31.45	0.9583	7.35	38.76	56.91	1.78	12.4	50	3.5
SRGAN	30.20	0.9456	6.21	35.44	53.89	2.32	22.9	80	5.2
SEPNet	32.10	0.9650	8.12	42.05	59.87	1.12	9.2	45	2.5

Original image (a) SRGAN (b) MICUNet

(c) CSHA (d) SEPNet (e) Improving CNN (Ours)

Figure 11: Image reconstruction effects of different algorithms.

Table 6: Statistical test results.

Index	Improved CNN vs CSHA	Improved CNN vs MICUNet	Improved CNN vs SRGAN	Improved CNN vs SEPNet
PSNR (dB)	t=4.32** d=1.25	t=5.67*** d=1.87	t=8.91*** d=3.02	t=2.15* d=0.63
SSIM	t=3.87** d=1.08	t=5.02*** d=1.52	t=9.24*** d=3.41	t=1.89 d=0.51
IFC	t=3.15** d=0.92	t=4.78*** d=1.63	t=7.53*** d=2.78	t=2.01* d=0.58
Inference time (ms)	t=6.45*** d=2.31	t=8.12*** d=3.15	t=12.6*** d=4.72	t=5.33*** d=1.94

Note: * p<0.05, ** p<0.01, *** p<0.001; Cohen's d>0.8 indicates a "large effect", and d>0.5 indicates a "medium effect". Statistical analysis is conducted using a two tailed independent samples t-test, α=0.05. Cohen's d is used to measure the size of the effect.

Table 7: Ablation results.

Component	PSNR (dB)	SSIM	IFC
(A) Baseline Model (SRCNN)	33.86	0.9068	4.79
(B) Main channel only (without guidance)	34.12	0.9135	4.88
(C) Main channel+guide channel (without mask)	34.35	0.9182	4.95
(D) Complete model (main+guide+mask)	34.97	0.9369	5.01
(E) Complete model (using standard optimizer)	34.68	0.9297	4.98

In Table 6, the improved CNN shows significantly better PSNR than all compared methods (p<0.01), especially for SRGAN (t=8.91, p<0.001). The difference in SSIM with SEPNet is not significant (p=0.062), but the PSNR is still significantly higher. The inference time effect of the above methods is generally large (d>1.9), and the difference in IFC between the improved CNN and CSHA is moderate (d=0.92), which verifies the role of mask guidance in information retention. Although the PSNR of SEPNet is slightly different from the improved CNN, its inference time is significantly longer, which does not meet real-time requirements. SRGAN is significantly inferior to the improved CNN method in all metrics, confirming that adversarial training is not suitable for scenarios that require high accuracy. To further evaluate the performance of each component in the research method, the ablation results are analyzed on the Set14 dataset, and the comparison results are shown in Table 7.

In Table 7, the improvement from (A) to (B) indicates that the improved model is superior to the original SRCNN, with a PSNR improvement of 0.26 dB (33.86→34.12). From (B) to (D), PSNR significantly increases by 0.85 dB (34.12→34.97), and SSIM increases by 0.0234 (0.9135→0.9369), demonstrating the core contribution of dual-channel mask-guided design. After introducing the mask, the PSNR increases by 0.62 dB again (34.35→34.97), indicating that providing explicit structural priors for guide channels is crucial. The improved optimizer can help the model converge to a better solution, with a PSNR value of 34.97 dB. This result effectively demonstrates the effectiveness of each component in the research method.

5 Discussion

Aiming at the problem of detail loss and efficiency bottleneck in super-resolution reconstruction of DM visual images, an improved CNN based on dual-channel guidance is proposed. The results showed that the dual-channel network effectively reduced the iteration loss value and significantly accelerated its convergence rate. The improved CNN had an accuracy value of over 95% on the dataset, with an MAPE value of less than 1.5%. The

MAPE values of 20 images using SC and KK were 3.92% and 4.12%, respectively, with lower response times compared to other comparative algorithms. The improved CNN had a 25% reduction in FLOPs, a better inference speed (20 ms) than SC (45 ms) and CNN (25 ms), and a PSNR value of 32.52 dB. In the image evaluation results, the PSNR and SSIM of the improved CNN on the CIFAR-10 dataset were only 0.15 dB and 0.0016 lower than SwinIR, and the MSSIM value was better than other methods. The research method maximally preserved the structural information of the original image, achieving the best information fidelity (8.47). The NQM value of the research method (43.98) performed well, increasing by 9.6% compared to CSHA (40.12), while SRGAN had the lowest NQM value (35.44) due to high-frequency artifacts introduced during adversarial training. Compared with SEPNet, the research method achieved reconstruction through one-time forward propagation, avoiding the time consumption of progressive methods. Compared with SRGAN, the research method outperformed in fidelity and was more suitable for DM applications that require high accuracy. The improved CNN exhibited better information fidelity (8.47) compared to other image reconstruction algorithms, and its WPSNR value (60.64 dB) verified the effectiveness of mask guidance. The significant improvement in IFC and NQM indicators indicated that the mask-guided channel effectively helped the model focus on restoring the structure and edges of the image, reducing blurring and artifacts, and meeting the needs of human visual perception. Although the improved model had slightly more parameters than the original SRCNN, its FLOPs (3.5G) were actually lower than some more complex models through a lightweight design and parallel computing of guide channels. This result achieved a good balance between performance and efficiency and had the potential for deployment on edge devices.

6 Conclusion

The generated DM visual images are basically consistent with real images, and have good image quality reconstruction effects and quality. The proposed lightweight architecture provides feasible technical solutions for DM application scenarios that are sensitive to computing resources, such as mobile image enhancement, real-time video super-resolution, and VR/AR content generation. Future research will focus on exploring more intelligent mask generation methods and trying to use neural structure search and meta-learning to automatically optimize the network structure. Dynamic pruning technology can be combined to reduce redundant parameters and computational burden and improve model adaptability. Moreover, this framework can be combined with attention mechanisms and multi-modal inputs to further enhance the robustness of reconstruction in complex scenes and improve the reconstruction ability of high-frequency details.

References

[1] R. Archana, and P. S. Eliahim Jeevaraj. Deep learning models for digital image processing: A review. Artificial Intelligence Review, 57(1):11, 2024. https://doi.org/10.1007/s10462-023-10631-z

[2] Gaurav Dhiman, A. Vignesh Kumar, R. Nirmalan, S. Sujitha, K. Srihari, N. Yuvaraj, P. Arulprakash, and R. Arshath Raja. Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. Multimedia Tools and Applications, 82(4):5343-5367, 2023. https://doi.org/10.1007/s11042-022-12178-7

[3] Shehzad Afzal, Sohaib Ghani, Mohamad Mazen Hittawe, Sheikh Faisal Rashid, Omar M. Knio, Markus Hadwiger, and Ibrahim Hoteit. Visualization and visual analytics approach for image and video datasets: A survey. ACM Transactions on Interactive Intelligent Systems, 13(1):1-41, 2023. https://doi.org/10.1145/3576935

[4] Tian Xie, and Kunpeng Zhao. SMPL variable model for 3D reconstruction and image fusion in animation media applications. IEEE Access, 13:93914-93929, 2025. https://doi.org/10.1109/ACCESS.2025.3549466

[5] Hee-jin Kim, Dong-seok Lee, and Soon-kak Kwon. Image reconstruction method by spatial feature prediction using CNN and attention. Journal of Multimedia Information System, 11(1):1-8, 2024. https://doi.org/10.33851/JMIS.2024.11.1.1

[6] Chao Tian, Kang Shen, Wende Dong, Fei Gao, Kun Wang, Jiao Li, Songde Liu, Ting Feng, Chengbo Liu, Changhui Li, Meng Yang, Sheng Wang, and Jie Tian. Image reconstruction from photoacoustic projections. Photonics Insights, 3(3):R06, 2024. https://doi.org/10.3788/PI.2024.R06

[7] Mahesh K. Singh, Sanjeev Kumar, Rajeev Ranjan, and Durgesh Nandan. Duplicate image detection using deep learning modified SVM and k-NN classification method for multimedia application. Soft Computing, 28(13):7659-7670, 2024. https://doi.org/10.1007/s00500-024-09756-2

[8] Miao Cao, Lishun Wang, Mingyu Zhu, and Xin Yuan. Hybrid CNN-transformer architecture for efficient large-scale video snapshot compressive imaging. International Journal of Computer Vision, 132(10):4521-4540, 2024. https://doi.org/10.1007/s11263-024-02101-y

[9] Yogesh Patel, Sudeep Tanwar, Pronaya Bhattacharya, Rajesh Gupta, Turki Alsuwian, Innocent Ewean Davidson, and Thokozile F. Mazibuko. An improved dense CNN architecture for deepfake image detection. IEEE Access, 11:22081-22095, 2023. https://doi.org/10.1109/ACCESS.2023.3251417

[10] Zhifeng Wang, Zhenghui Wang, Chunyan Zeng, Yan Yu, and Xiangkui Wan. High-quality image compressed sensing and reconstruction with multi-scale dilated convolutional neural network. Circuits,

Systems, and Signal Processing, 42(3):1593-1616, 2023. https://doi.org/10.1007/s00034-022-02181-6

[11] Yuantao Chen, Runlong Xia, Kai Yang, and Ke Zou. DGCA: High resolution image inpainting via DR-GAN and contextual attention. Multimedia Tools and Applications, 82(30):47751-47771, 2023. https://doi.org/10.1007/s11042-023-15313-0

[12] Tomoya Nakamura, Takuto Watanabe, Shunsuke Igarashi, Xiao Chen, Kazuyuki Tajima, Keita Yamaguchi, Takeshi Shimano, and Masahiro Yamaguchi. Superresolved image reconstruction in FZA lensless camera by color-channel synthesis. Optics Express, 28(26):39137-39155, 2020. https://doi.org/10.1364/OE.410210

[13] Xiuxi Pan, Xiao Chen, Saori Takeyama, and Masahiro Yamaguchi. Image reconstruction with transformer for mask-based lensless imaging. Optics Letters, 47(7):1843-1846, 2022. https://doi.org/10.1364/OL.455378

[14] Yuantao Chen, Linwu Liu, Phonevilay Volachith, Ke Gu, Runlong Xia, Jingbo Xie, Qian Zhang, and Kai Yang. Image super-resolution reconstruction based on feature map attention mechanism. Applied Intelligence, 51(7):4367-4380, 2021. https://doi.org/10.1007/s10489-020-02116-1

[15] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, and Vishal Goyal. Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion, 91:230-260, 2023. https://doi.org/10.1016/j.inffus.2022.10.007

[16] Yang Yang, Chang Liu, Hui Wu, and Dingguo Yu. A distorted-image quality assessment algorithm based on a sparse structure and subjective perception. Mathematics, 12(16):2531, 2024. https://doi.org/10.3390/math12162531

[17] Liming Zhang, Heng Li, Rui Zhu, and Ping Du. An infrared and visible image fusion algorithm based on ResNet-152. Multimedia Tools and Applications, 81(7):9277-9287, 2022. https://doi.org/10.1007/s11042-021-11549-w

[18] Guifang Shao, Qiao Sun, Yunlong Gao, Qingyuan Zhu, Fengqiang Gao, and Junfa Zhang. Sub-pixel convolutional neural network for image super-resolution reconstruction. Electronics, 12(17):3572, 2023. https://doi.org/10.3390/electronics12173572

[19] Zhi Jin, Muhammad Zafar Iqbal, Dmytro Bobkov, Wenbin Zou, Xia Li, and Eckehard Steinbach. A flexible deep CNN framework for image restoration. IEEE Transactions on Multimedia, 22(4):1055-1068, 2020. https://doi.org/10.1109/TMM.2019.2938340

[20] Yu Zhang, Yu Liu, Peng Sun, Han Yan, Xiaolin Zhao, and Li Zhang. IFCNN: A general image fusion framework based on convolutional neural network. Information Fusion, 54(1):99-118, 2020. https://doi.org/10.1016/j.inffus.2019.07.011

[21] Haiying Xia, Fuyu Zhu, Haisheng Li, Shuxiang Song, and Xiangwei Mou. The combination of multi-scale and residual learning in deep CNN for image denoising. IET Image Processing, 14(10):2013-2019, 2020. https://doi.org/10.1049/iet-ipr.2019.1386

[22] Shuping Yuan, Yang Chen, Chengqiong Ye, and Mohd Dilshad Ansari. Edge detection using nonlinear structure tensor. Nonlinear Engineering, 11(1):331-338, 2022. https://doi.org/10.1515/nleng-2022-0038

[23] Mallika Garg, Jagpal Singh Ubhi, and Ashwani Kumar Aggarwal. Neural style transfer for image steganography and destylization with supervised image to image translation. Multimedia Tools and Applications, 82(4):6271-6288, 2023. https://doi.org/10.1007/s11042-022-13596-3

[24] Chunsheng Wei, Qifeng Li, Xiaodong Zhang, Xiangyun Ma, and Jianbin Du. A fast snapshot hyperspectral image reconstruction method based on three-dimensional low rank constraint. Canadian Journal of Remote Sensing, 47(4):588-595, 2021. https://doi.org/10.1080/07038992.2021.1943340

[25] Doo Bin Kim, Hyun-Mee Park, Sang-Hoon Joon, and JooWan Hong. Effectiveness of noise reduction in LDCT images based on SRCNN. Korean Journal of Artificial Intelligence, 12(4):21-26, 2024.

[26] Ming Yu, Jiecong Shi, Cuihong Xue, Xiaoke Hao, and Gang Yan. A review of single image super-resolution reconstruction based on deep learning. Multimedia Tools and Applications, 83(18):55921-55962, 2024. https://doi.org/10.1007/s11042-023-17660-4