Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105
Volume **11**, Number **1**, February **2026**

A Visual Communication Design Approach Based on Wavelet Transform and Deep Learning

Jia-Ying Cai[1]

[1]College of Design and Art,
Liuzhou Institute of Technology, Liuzhou 545000, P. R. China
titiying0814@163.com

Kai Liu[2,*]

[2]College of Humanities, Arts and Design,
Guangxi University of Science and Technology, Liuzhou 545000, P. R. China
liukai_6776@163.com

Rui-Hui Wu[3]

[3]Faculty of Entrepreneurship and Business,
Universiti Malaysia Kelantan, Kuantan 16100, Malaysia
a20e0276f@siswa.umk.edu.my

[*]Corresponding author: Kai Liu

Received November 06, 2024, revised February 25, 2025, accepted May 24, 2025.

Abstract. *Existing visual communication design approaches are accompanied by noise and distortion, with blurred image detail information. To address these problems, this article suggests a visual communication design approach relied on wavelet transform (WT) and deep learning. The wavelet VisuShrink threshold and soft threshold function of WT are first improved, and the image is denoised using the improved WT, which has good structure preservation in the process of image multilayer wavelet decomposition and reconstruction denoising, and retains the image texture details while removing the noise effectively. Then a multistage null CNN is introduced to capture the deep characteristics of the denoised picture. Through the hybrid attention module (HAM), the attention map is generated from both channel and spatial dimensions, OWT is introduced into HAM to reduce the information loss of the spatial information compression function, and the model autonomously learns the donation of every characteristic channel to the completion of the visual communication task, and gives more attention to the ones with large contribution, so optimizing the visual communication impact. Experimental outcome on the publicly available dataset MIT-Adobe5K indicates that the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the offered approach are 24.2dB and 0.9461, respectively, which can remove the noise and also enhance the image details well, and the enhanced image is more in line with the human eye's visual effect.*
Keywords: Visual communication; Wavelet transform; Deep learning; Null CNN; Hybrid attention mechanism.

1. **Introduction.** Vision is the chief way for human beings to acquire information, and among the information acquired visually, image carriers occupy a high proportion [1]. Compared with text, voice and other carriers, images are more intuitive and carry more information, so they are widely used in daily information exchange [2]. However, in many real-life scenarios, the captured images are featured by low brightness, low contrast,

narrow gray scale range, and are accompanied by noise, distortion and other phenomena, which affect the subjective visual impact of the human eye and intuitively obtaining information from the images, and to a certain extent, affect the accurate and timely delivery of information, so it is essential to enhance the images [3]. Visual communication refers to the communication of relevant content through visual media [4], the technology can show the design sense of graphics and rich information, due to the continuous growth of digital technique, the visual communication technology has been greatly challenged, so it is essential to visually improve the picture for the goal of strengthening the effect of visual communication.

1.1. **Related work.** To enhance the visual communication, Ulutas and Ustubioglu [5] used Adaptive Histogram Equalization (AHE) algorithm for optimization of visual communication, the enhanced image edges and texture details are better preserved and strengthened, but the image has more noise. To overcome the problem that the AHE algorithm can over-enhance the noise, Paul [6] suggested a contrast-limited AHE (CLAHE) algorithm, which crops and redistributes the portion of the histogram that exceeds a threshold to other gray levels, which enhances the contrast and reduces the noise generation at the same time. Yamakawa and Sugita [7] offered an improved picture enhancement approach relied on the Retinex approach, but it is not effective in enhancing the contours and details of the objects in the image. Subramani and Veluchamy [8] achieved good results in image enhancement by using improved color correction, adaptive lookup table, and fast localized Laplacian filter. The enhancement results are good.

Compared to the traditional visual communication enhancement methods mentioned above, the wavelet transform (WT) adopts a set of wavelet basic functions to represent or approximate the signal, which not only improves the clarity and contrast of the image, but also effectively controls the noise and makes the image contours more visible. Ren et al. [9] processed the wavelet coefficients using multiscale nonlinear high pass filtering for image enhancement. Muthukrishnan et al. [10] designed a picture improvement approach using WT relied on the inflection operation and gamma correction, which applies the inverse WT to gain an optimized picture. Laksmi et al. [11] proposed to convert the RGB of the image to HSV channel and at the same time do discrete wavelet transform of the luminance channel to isolate frequency subbands, and also use fuzzy transform to achieve the improvement and denoising of the detailed information.

However, WT-based visual communication improvement methods require extensive computation and manual intervention for the selection of transformation parameters, it is the breathtaking performance of deep learning that has revolutionized visual communication [12, 13]. Compared to traditional methods, Deep Learning (DL) has the advantage of higher accuracy, better robustness and faster speed. Ni et al. [14] designed a GAN based image enhancement structure which utilizes the powerful ability of GAN networks to produce pictures from real data distributions and the outcome proved its efficiency. Wan et al. [15] utilized the advantages of GAN and UNET and proposed the "Enlighten GAN" network to obtain excellent image results using unpaired images for training. GAN-based visual enhancement methods are inadequate for feature extraction, while convolutional neural networks (CNN) show more power in feature representation and nonlinear mapping [16]. Kuang et al. [17] suggested a CNN-based image enhancement method and added an attention mechanism to the network framework for acquiring richer features, which improves the visual effect while recovering the color cast and blurriness of the image. Singh and Parihar [18] designed a null convolution-based top-bottom attention model, which utilizes convolution kernels with different null rates to capture both local and global semantic information of a picture, and enhances the visual effect of the image.

1.2. **Contribution.** To summarize, existing visual communication studies have noisy images and unclear detail information, resulting in poor visual communication. Therefore, to cope with the above issues, this article suggests a visual communication design method based on wavelet transform and deep learning. The innovative work of the method is reflected in the following aspects.

(1) An improved WT algorithm (OWT) is offered, based on VisuShrink thresholding to improve the thresholding, which can better separate the noise and improve the denoising effect during wavelet processing; in addition, the algorithm is improved based on the soft thresholding function, which overcomes the disadvantage of the transition of the soft thresholding function that is not smooth and makes the denoised image more natural.
(2) A multilevel null CNN is proposed to extract image features. The structure expands the receptive field through multi-level cavity convolution and utilizes a large range of semantic information to guide the recovery of dark regions of the image. A hybrid attention module (HAM) is introduced to generate an attention map from both channel and spatial dimensions to ensure that the enhanced image is structurally complete and texturally clear.
(3) Replace the average pooling and maximum pooling in HAM with OWT to reduce the information loss of the spatial information compression function, and autonomously learn the contribution of every characteristic channel to accomplish the enhancement task, giving more attention to the features with high contribution and ignoring the redundant or low contribution features, so as to enhance the visual communication effect.
(4) Simulation experiments were carried out on the open data set MIT-Adobe5K, and the outcome implied that the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the proposed model increased by 4.08% to 25.69% compared with other models, which not only has a strong image denoising ability, but also can effectively enhance the clarity of images. It has good application prospect in the task of visual communication design.

2. Theoretical analysis.

2.1. **Wavelet transforms.** WT is a mathematical way for digital image processing that can process images in the time-frequency domain and can achieve high compression ratios while retaining the important visual features of the image [19], as shown in Figure 1. WT provides a frequency-domain transform that can be varied in scale and shifted on the time axis. The wavelet function is defined as follows.

$$WT(\alpha,\tau)=\frac{1}{\sqrt{\alpha}}\int_{-\infty}^{+\infty}f(t)\psi^{*}\left(\frac{t-\tau}{\alpha}\right)dt \tag{1}$$

Figure 1. Schematic diagram of wavelet transform

where $f(t)$ is the analyzed signal, $\psi(t)$ is the wavelet mother function, $\psi^{*}(t)$ is the conjugate complex of $\psi(t)$, α is the scale function, and τ is the translation distance controlling the translation of the wavelet function.

WT uses a set of basis functions, and different wavelet basis functions represent different transformation results. Common wavelet bases are haar, db, sym, bior, etc [20],

these wavelet bases are localized and multi-scale in nature, which can decompose and reconstruct the signal at different times and frequencies to preserve the temporal or spatial information [20, 21]. The wavelet basic function is defined as follows.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

where $a, b \in R$.

In visual communication, WT is commonly used for image compression and denoising. The fundamental idea of compression is to decrease the amount of data in an image by removing high-frequency wavelet coefficients, thus realizing picture compression. The fundamental idea of denoising is to disintegrate the picture into wavelet coefficients of different scales, and reduce the effect of noise by removing the high-frequency wavelet coefficients, thus realizing the denoising of the image. In this paper, WT will be utilized for denoising and compression of images to enhance the visual communication.

2.2. **Convolutional neural network.** CNN is a ordinary DL model, compared with other neural networks, its advantage is to reduce the amount of weights can make the network simple to improve, and at the same time decreases the model complexity and reduces the hazard of overfitting. CNN mainly includes convolutional level, pooling level, and full connectivity level [22].

(1) A convolutional level usually consists of several convolutional kernels. The number of convolution kernels reflects the amount of channels of the characteristic picture generated by the convolution layer. When there are multiple convolutional levels and multiple convolutional operations, the formula for calculating the input size and output size of an image after a convolutional level is as follows.

$$Out = (Input + 2 \times Padding - kernel - (kernel - 1) * (d - 1))/stride + 1 \tag{3}$$

where $Input$ is the length or width of the input characteristic map, $Padding$ is the input padding width of the current convolutional level, $kernel$ is the size used, d is the void factor of the convolutional level with a default value of 1, $stride$ is the step size of the convolutional kernel, and Out is the size of the final output characteristic map.

(2) The pooling level can count an area in the characteristic map, extract the feature values that best represent the area, and abstract the area using a single result. The process of average pooling and max pooling is indicated in Figure 2.

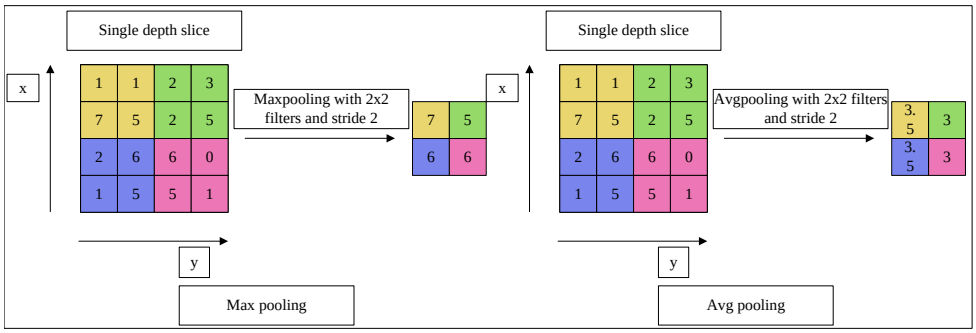

Figure 2. The process of average pooling and max pooling

(3) The fully connected level calculates the score of each class based on the extracted characteristic values and maps the "distributed feature representation" of the feature space to the sample labeling space. Since the fully connected level is computationally expensive, the global pooling level and the average pooling level have been added to efficiently reduce the parameters in the network.

3. Image denoising based on optimized wavelet transforms.

3.1. Wavelet transform optimization based on optimized VisuShrink thresholding and soft thresholding functions.

A picture denoising approach relied on optimized wavelet transform (OWT) is proposed for the phenomenon that visual images are often accompanied by noise. The algorithm improves the wavelet VisuShrink threshold and soft threshold function. The WT algorithm with both improvements applied simultaneously can fruitfully remove the noise while preserving the texture details of the image.

In the process of WT image denoising, the selection of threshold is an important part. The improved VisuShrink threshold with the introduction of adaptive parameters has good structure preservation in the process of image multilayer wavelet decomposition reconstruction and denoising; the improved soft threshold function with the addition of the transition of the smoothing function is smoother in the part of the low wavelet coefficients and is able to eliminate the effect of the intermittent points.

(1) **VisuShrink threshold improvement.** As the VisuShrink threshold, which is widely used and has strong correlation with images, has been improved and used by many scholars [23], so in this paper, the VisuShrink threshold is selected for improvement. Unimproved VisuShrink thresholding tends to yield large thresholds when the image is large, which tends to corrupt the image information. The improved VisuShrink thresholds in this chapter retain the scale weights of the VisuShrink thresholds and increase the importance of the standard deviation estimates, while overcoming the disadvantages of over-processing of the VisuShrink thresholds and the lack of adaptivity of the fixed scale thresholds. The improved VisuShrink threshold is shown below:

$$\lambda = \frac{\sigma^2\sqrt{2\ln N}}{MAX} \tag{4}$$

where σ is the standard deviation of the noise, N is the amount of signals or pixels in the picture and MAX is the maximum value of the wavelet coefficients.

(2) **Threshold function improvement.** The commonly used thresholding functions in WT denoising are soft thresholding function, hard thresholding function and garrote thresholding operation [24]. The denoising effect of soft threshold operation is often better, for the noise image constant deviation can be more effective in suppressing the noise while maintaining the loss of information in a certain controllable range, so this paper chooses soft threshold function as the basis for improvement. To overcome the disadvantage of soft threshold function in processing wavelet coefficients near the threshold point which are not smooth, a smooth curve is used to replace the processing at the end of low wavelet coefficients, and the characteristics of soft threshold function are retained at larger wavelet coefficients. This is implied below:

$$\hat{w}_{i,j} = \begin{cases} sgn(w_{i,j})(|w_{i,j}| - \lambda), & |w_{i,j}| \geq 2\lambda \\ sgn(w_{i,j})\frac{w_{i,j}^2}{4}, & |w_{i,j}| < 2\lambda \end{cases} \tag{5}$$

3.2. **Visual image denoising based on optimized wavelet transforms.** After improving the VisuShrink threshold and soft threshold function of WT, the visual images are denoised using OWT. Assuming that the image pixel is $x(i,j)$, which is regarded as the extreme value of the approximation pixel, each level of the approximation pixel is smoothed by the low-pass filtering function $\varphi(x)$. When there are many small textures in the high-frequency component, the WT can also decompose the high-frequency information arbitrarily, and the decomposition expression of the WT subspace V_j is as follows, in which j is a constant term:

$$\begin{cases} V_j = U_{j-1}^2 \oplus U_{j-1}^3 \\ \cdots \\ V_j = U_{j-k}^{2^k} \oplus U_{j-k}^{2^k+1} \oplus \cdots \oplus U_{j-k}^{2^{k+1}-1} \\ \cdots \\ V_j = U_0^{2^j} \oplus U_0^{2^j+1} \oplus \cdots \oplus U_0^{2^{j+1}-1} \end{cases} \tag{6}$$

Suppose γ_j is the scale subspace, where, WT package $\psi_{j,k,n}(t)$ is as follows:

$$\begin{cases} \psi_{J,K,N'}(t) = 2^{-\frac{J}{2}} \psi_n(2^{-J}t - k) \\ \psi_n(t) = 2^{-\frac{J}{2}} \psi_{u2^l+m}(2^l t) \end{cases} \tag{7}$$

where J is the scale parameter, K is the position parameter, N' is the frequency parameter, and u is the parameter required for OWT.

Due to the existence of frequency parameters in the wavelet packet, the disadvantage of low resolution in the process of wavelet resolution is solved. The image denoising based on the improved VisuShrink threshold is to collect signals through the wavelet packet. When the absolute value of the wavelet packet coefficient is less than the threshold value λ, the wavelet packet coefficient is noise; otherwise. Finally, thresholds are set for high and low frequencies respectively to obtain an excellent de-noised image X.

4. Visual communication enhancement based on wavelet transform and deep learning.

4.1. **Null CNN-based visual image feature extraction.** After wavelet denoising the image, to solve the problems such as blurring of picture detail data in the existing visual communication design methods, firstly, a multilevel null CNN is designed to capture the depth characteristics of the picture, and then the HAM is used to generate the attention picture from the channel and spatial dimensions, and OWT is introduced into the HAM to reduce the information loss of the spatial information compression function, and the model autonomously learns that each feature channel contributes to the completion of the visual communication task and ignores redundant or less contributing features to optimize the visual communication effect. The model structure of the designed approach is indicated in Figure 3.

CNN can effectively mine image features by stacking multiple convolutional layers. However, this approach needs a large amount of training data to prevent overfitting during training. To address this problem, Wiatowski and Bölcskei [25] utilized cavity convolution to obtain a large range of feature information in the input feature map, which improves the adaptability of convolutional neural networks. Based on this, in this paper, a multilayer self-conducting cavity dense structure is designed to capture the global features in the picture, as implied in Figure 4.

A hollow convolution with a large void coefficient is used to guide a hollow convolution with a small void coefficient from the bottom up. This allows the normal convolution to

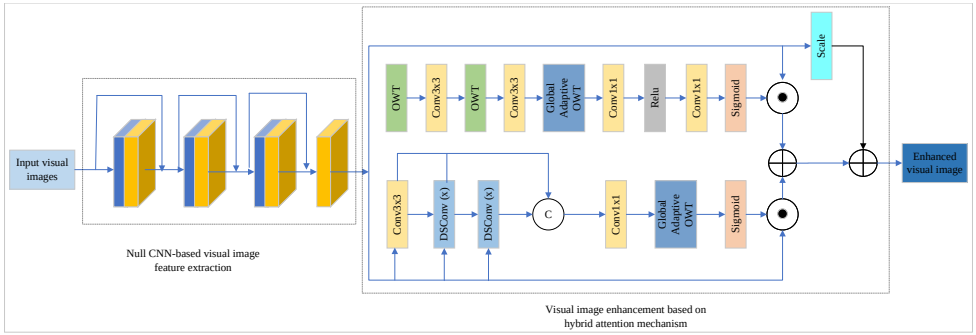

Figure 3. The model structure of the designed approach

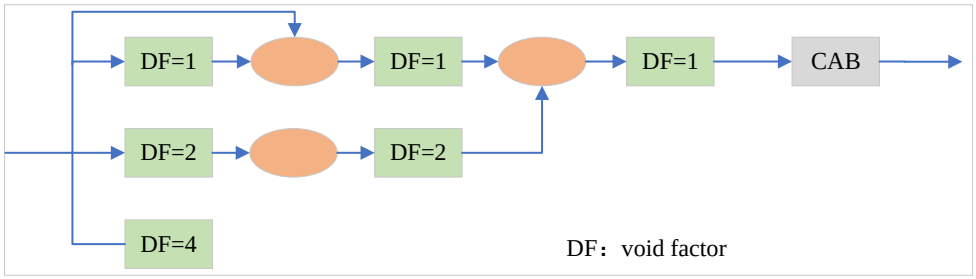

Figure 4. The multilayer self-conducting cavity dense structure

have a larger sensory field and accurately extract the global information in the image. Specifically, the structure utilizes multilayer null convolution to obtain feature information with different receptive field sizes from the input feature map. Its corresponding expression is as follows:

$$y_r = Conv_r(X) \tag{8}$$

where $r = 1, 2, \ldots, 2^{L-1}$, X is the visual image processed by WT denoising, y is the output feature map, $Conv_r$ is the null convolution with null coefficient r, and L is the number of layer levels of the multilevel self-guided null dense block. This structure performs guided learning between two neighboring layers, using the layer with larger sensory field to guide the layer with smaller sensory field for image feature extraction. The corresponding expression is as follows:

$$l_r = \begin{cases} y_r, & r = 2^{L-1} \\ Conv_r(cat[y_r, y_{r/2}]), & r = 2, \ldots, 2^{L-2} \\ Conv_1(cat[y_1, X]), & r = 1 \end{cases} \tag{9}$$

where l_r is the final output of each level, cat is the splicing operation of the feature map, three null convolutions with different null coefficients extract the image information and then fused on the main line branch to get the final feature map as Z.

$$Z = Concat(l_1, l_2, \ldots, l_r) \tag{10}$$

4.2. **Visual image enhancement based on hybrid attention mechanism.** After obtaining the fused feature maps, this paper utilizes the hybrid attention mechanism to modify the weights of the feature maps at various levels. To be able to highlight the global and local characteristics of the feature map, this paper uses a hybrid attention module (HAM) combining global consistent channel attention (GCCA) and local topological spatial attention (LTSA) to focus on the detail information of the visual image.

HAM consists of a global consistent channel attention branch and a local topological space attention branch, which are described below.

(1) Consistent channeled attention across the board. Assuming that Z has l feature maps of size $H \times W$, and l is the number of channels input to Z. Unlike the existing Channel Attention Module (CAM) which only performs one downsampling operation, the GCCA firstly compresses the image of Z through the $AvgPool$ operation and the $MaxPool$ operation, to decrease the loss of information brought by the one downsampling operation, with the following equation:

$$F_{gf} = Conv3 \times 3(MaxPool(Conv3 \times 3(AvgPool(X)))) \tag{11}$$

The global information is extracted by reducing the spatial dimension of the input characteristic picture to 1 using global adaptive mean pooling, which is formulated as follows:

$$F'_{gf} = GlobalAdaptiveAvgPool(F_{gf}) \tag{12}$$

Finally, the channel features with global attention are obtained by a convolutional layer with a convolutional kernel size of 1×1, $ReLU$ activation function and $Sigmoid$ activation function, which are formulated as follows:

$$F_{GCCA} = Sigmoid(Conv1 \times 1(ReLU(Conv1 \times 1(F'_{gf})))) \tag{13}$$

(2) Local topological spatial attention. Assuming that Z is a feature map of size $H \times W \times C$, it is subjected to a conventional convolution with a convolution kernel size of 3 and two dynamic serpentine convolutions on the x and y axes, respectively. While each convolution kernel of traditional convolution can only perceive a fixed position of the input, dynamic serpentine convolution [26] is based on a dynamic structure that is better adapted to the detailed texture structure of the picture to better perceive the key features of the picture, and its principle equation is as follows:

$$F_{DS_x} = DSConv(Z_x) \tag{14}$$

$$F_{DS_y} = DSConv(Z_y) \tag{15}$$

where F_{DS_x} and F_{DS_y} are x-axis shifted feature maps and y-axis shifted feature maps respectively. Next, connect y_r with F_{DS_x} and F_{DS_y} as shown below:

$$F_{concat} = Concat(y_r, F_{DS_x}, F_{DS_y}) \tag{16}$$

The connected characteristic maps are then fed into a convolutional level with a convolutional kernel size of 1. The channel dimensions are downscaled by an adaptive mean pooling level and a sigmoid activation function level, and the LTSA is obtained as follows:

$$F_{LTSA} = Sigmoid(adaptiveAvgPool(Conv1 \times 1(F_{concat}))) \tag{17}$$

The resulting GCCA and LTSA are multiplied with the input feature Z at the element level respectively, and the two feature mappings are fused using addition to obtain the HAM features. In the jump join, this method uses a learnable scale factor $s \in R^1$ for better performance.

$$F_{Mix} = F_{LTSA} \cdot Z + F_{GCCA} \cdot Z + s \cdot Z \tag{18}$$

4.3. **Visual communication optimization based on wavelet transform and deep learning.** In the previous paper, feature extraction and enhancement of visual images were achieved by CNN and HAM respectively, but the use of $AvgPool$ and $MaxPool$ together as the image compression function in HAM may bring about a large loss of detail information [27], for this reason, in this paper, the optimized WT is used instead of $AvgPool$ and $MaxPool$ operations in HAM for image compression. Given a visual image X, OWT uses four filters, namely low-pass filters g_{LL} and high-pass filters g_{LH}, g_{HL}, and g_{HH}, to decompose X into four subband images x_{LL}, x_{LH}, x_{HL}, and x_{HH}. WT has different forms according to different filters selected. In this paper, the most commonly used Haar wavelet is selected, and the convolution step of the four Haar wavelet filters is set to 2 during the convolution process, as shown below:

$$g_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, g_{LH} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, g_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, g_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{19}$$

Thus the four components after OWT can be obtained from Equation (20):

$$\begin{cases} x_{LL}(i,j) = x(2i-1, 2j-1) + x(2i-1, 2j) + x(2i, 2j-1) + x(2i, 2j) \\ x_{LH}(i,j) = -x(2i-1, 2j-1) - x(2i-1, 2j) + x(2i, 2j-1) + x(2i, 2j) \\ x_{HL}(i,j) = -x(2i-1, 2j-1) + x(2i-1, 2j) - x(2i, 2j-1) + x(2i, 2j) \\ x_{HH}(i,j) = x(2i-1, 2j-1) - x(2i-1, 2j) - x(2i, 2j-1) + x(2i, 2j) \end{cases} \tag{20}$$

Haar wavelet decomposes the image into four images, and the spatial compression based on this will retain more details than the information compression using global average pooling and global maximum pooling on the original characteristic map, so in this paper, the input characteristic map is first decomposed into four components by OWT. To further accomplish the information integration, the components generated by Haar wavelet were further integrated with information using a null CNN before constructing the HAM. The fully-connected level projects the integrated information into a tighter space with a ratio r. Converting l level of input into l/r level reduces both computational effort and feature redundancy. Subsequently, the features are recovered from the space and the l/r layers are remapped to l levels. Finally, after redistributing the weights of the levels, HAM is used to enhance the detail information and optimize the visual communication.

5. Experiment and result analysis.

5.1. **Quantitative analysis of experimental results.** The dataset used in the experiments is the MIT-Adobe5K dataset [28], a commonly used dataset in the field of visual communication, which is provided by Adobe, and contains 5000 images covering natural landscapes, buildings, and portraits of people. In this article, the dataset is classified into training set and test set according to 7:3. The CPU used for training is Intel Xeon E5-2620 v4, the GPU is NVIDIA Tesla K80, and the network algorithm is built in the torch environment of version 1.8.1. The Adam optimizer is responsible for optimization during the network training process, the number of training iterations Epoch is 1000, the number of images used for training in each batch is 50, the starting studying rate is 0.0001, and the learning rate gradient descent degree is 0.7.

To estimate the effectiveness of the designed method, the proposed method OURS is compared and experimented with the WTEN method in literature [10], the GANET method in literature [15], the CNNAM method in literature [17] and the NUAM method in literature [18]. Firstly, the test time consuming and training time consuming for visual information optimization with different number of images are investigated, and the

experimental outcome is implied in Figure 5. The test time consuming and training time consuming of OURS are lower than other methods. When the number of images is 100, the test elapsed time and training elapsed time of OURs are 5.1s and 11.8s, respectively, which are 13.8s and 20.9s lower than WTEN, 10.6s and 13.1s lower than GANET, 9.1s and 11s lower than CNNAM, and 3.4s and 4.1s lower than NUAM. it shows that the WT- and DL-based OURS methods are computationally efficient and can efficiently achieve the optimization of image visual communication.

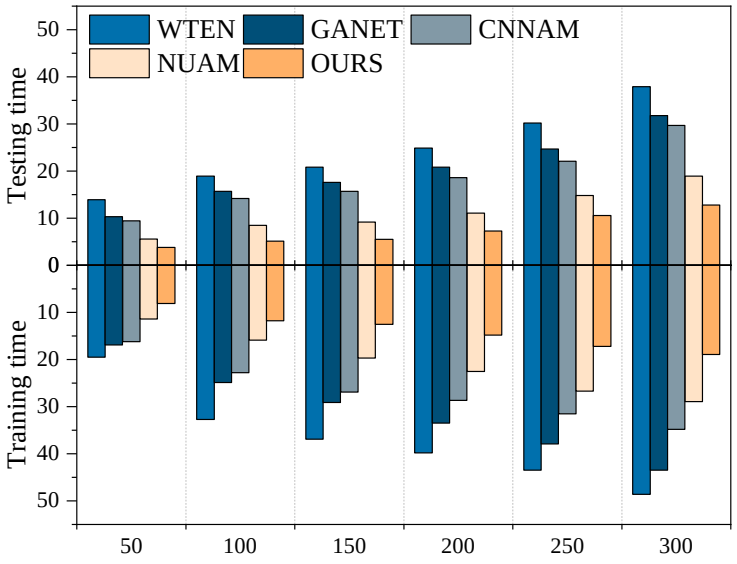

Figure 5. Testing time-consuming and training time-consuming visual communication design methods

The total elapsed time to optimize the visual communication effect of different methods is formed into a log file during the experiment and then run through Python to generate a box-and-whisker plot, as shown in Figure 6. The total elapsed time of OURS does not fluctuate up and down by more than 2.1s, with a maximum of only 15.2s, while the elapsed time of WTEN fluctuates the most, with a maximum elapsed time of 31.9s. The elapsed time of NUAM is the next best performance, with a maximum elapsed time of only 19.3s. The longest elapsed time for GANET and CNNAM is 28.9s and 26.8s, respectively. Therefore, it can be observed that OURS has a more centralized distribution of values than the other methods, which indicates that OURS is more stable.

5.2. **Performance analysis of visual communication effects.** To objectively assess the effect of visual communication, this paper evaluates the image enhancement effect of different methods using PSNR, SSIM, Image Quality Evaluation (NIQE), and Visual Fidelity (VIF), which are commonly used metrics to measure the effect of visual communication. Figure 7 shows the comparison of PSNR of different methods when the number of iterations is from 0 to 500. PSNR is used to evaluate the difference in quality between the initial picture and the distorted image. Higher value of PSNR indicates better quality of image and lesser distortion. The PSNR metric of OURS is higher than the other four models, with 16.9 dB, 17.5 dB, 21.2 dB, 22.6 dB, and 23.8 dB for WTEN, GANET,

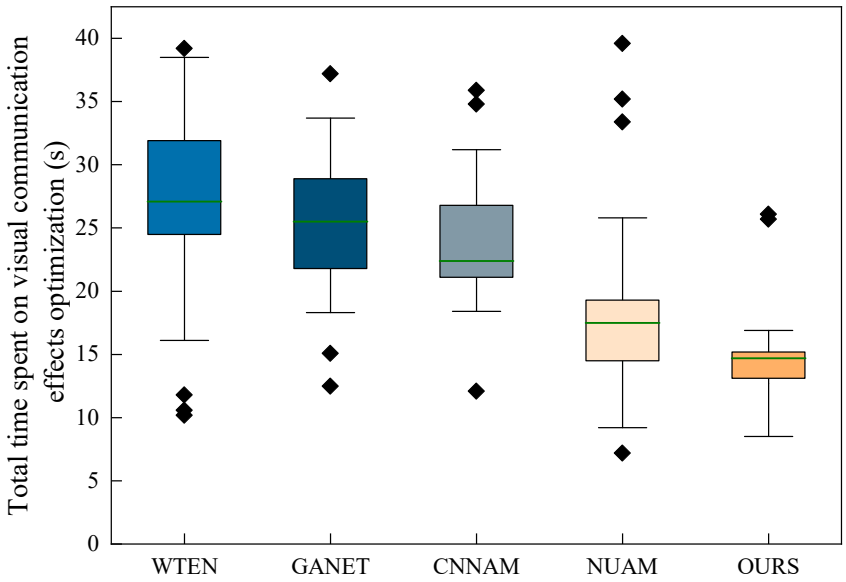

Figure 6. Total time-consuming box-and-whisker diagrams for visual communication design methods

CNNAM, NUAM, and OURS, respectively, when Epochs are 300. OURS denoises the image using OWT and also compresses the image using OWT to retain the maximum possible detail information without distortion, and enhances the key features using HAM to improve the visual communication.

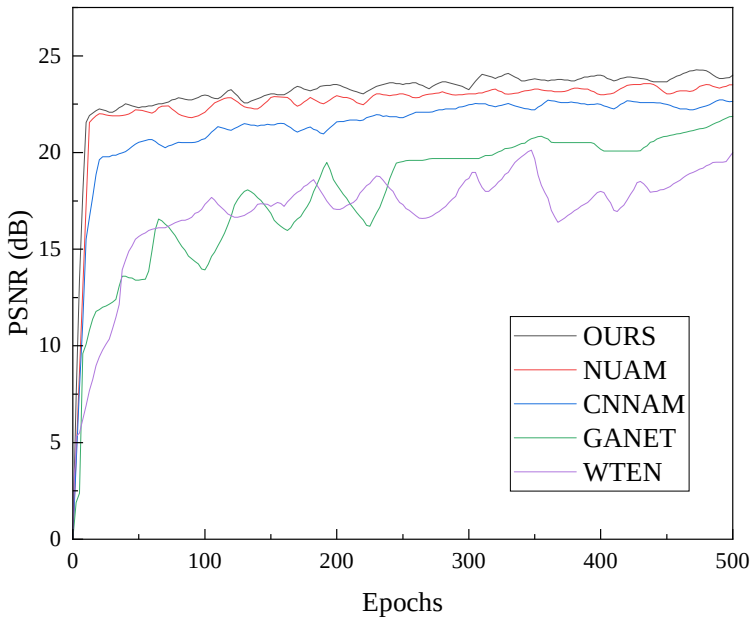

Figure 7. PSNR of different visual communication design methods

Comparisons of SSIM, NIQE, and VIF for the different methods are shown in Table 1, where ↑ indicates that the higher the value the better, ↓ indicates that the lower the value the better, and the bolded black font indicates the best result in this comparison.

As shown in Table 1, OURS improves 5.24%-25.69% on SSIM and VIF compared to existing methods. 3.17%-8.79% reduction in NIQE compared to the existing methods. WTEN achieves image enhancement only through WT, which significantly increases the computational effort due to the need for multi-layer decomposition and reconstruction operations. GANET does not extract image features sufficiently and does not highlight enough details, resulting in poor visualization. Both CNNAM and NUAM are based on CNN for image enhancement, but they do not consider image denoising, so the visual communication effect is not as good as that of OURS. In summary, OURS-enhanced image has a better visual communication effect.

Table 1. Comparison of performance indicators for visual communication effects

Method	SSIM↑	NIQE↓	VIF↑
WTEN	0.7842	4.8784	0.7362
GANET	0.8124	4.4968	0.7851
CNNAM	0.8762	4.1193	0.8496
NUAM	0.8927	3.6481	0.8793
OURS	**0.9461**	**3.3272**	**0.9254**

6. **Conclusion.** Visual communication enhances the visual effect by enhancing the image brightness and recovering the image content. Aiming at the problems of unclear image details and poor visual communication effect of the existing visual communication design approaches, this paper firstly improves the VisuShrink threshold and threshold function in WT, which can better separate the noise in the process of wavelet processing and make the denoised image more natural. Then a multilevel null CNN is designed to extract image features, expanding the perceptual field through multilevel null convolution, introducing HAM to generate attention maps from both channel and spatial dimensions, replacing average pooling and maximum pooling in HAM with OWT to reduce the information loss of the spatial information compression function, and autonomously learning the contribution of each characteristic channel to the completion of the enhancement task, so as to enhance the visual communication effect. The experimental results show that the proposed model has high PSNR and SSIM, and can effectively realize the enhancement of visual communication effect. Through theoretical research and experiments, it is proved that the method studied in this paper has achieved certain effect, but the experimental dataset is relatively single, and in the future, comparative experiments will be carried out on multiple visual images to verify the generalization performance of the design model.

Acknowledgment. This work is supported by the Path Exploring of Ideological and Political Construction in the "Design Thinking" Course under the Background of Smart Education (No. 2023JGB504), and the Research on Display Design Strategies for the Fusion of Guangxi Intangible Cultural Heritage with Metaverse in Diverse Contexts (No. 23BXW002).

REFERENCES

[1] Z. Tian, "Dynamic visual communication image framing of graphic design in a virtual reality environment," *IEEE Access*, vol. 8, pp. 211091–211103, 2020.

[2] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.

[3] K.-F. Yang, X.-S. Zhang, and Y.-J. Li, "A biological vision inspired framework for image enhancement in poor visibility conditions," *IEEE Transactions on Image Processing*, vol. 29, pp. 1493–1506, 2019.
[4] U. Russmann, and J. Svensson, "Introduction to visual communication in the age of social media: Conceptual, theoretical and methodological challenges," *Media and Communication*, vol. 5, no. 4, pp. 1–5, 2017.
[5] G. Ulutas, and B. Ustubioglu, "Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation," *Multimedia Tools and Applications*, vol. 80, pp. 15067–15091, 2021.
[6] A. Paul, "Adaptive tri-plateau limit tri-histogram equalization algorithm for digital image enhancement," *The Visual Computer*, vol. 39, no. 1, pp. 297–318, 2023.
[7] M. Yamakawa, and Y. Sugita, "Image enhancement using Retinex and image fusion techniques," *Electronics and Communications in Japan*, vol. 101, no. 8, pp. 52–63, 2018.
[8] B. Subramani, and M. Veluchamy, "Quadrant dynamic clipped histogram equalization with gamma correction for color image enhancement," *Color Research & Application*, vol. 45, no. 4, pp. 644–655, 2020.
[9] Z. Ren, G. Ren, and D. Wu, "Fusion of infrared and visible images based on discrete cosine wavelet transform and high pass filter," *Soft Computing*, vol. 27, no. 18, pp. 13583–13594, 2023.
[10] A. Muthukrishnan, D. V. Kumar, and M. Kanagaraj, "Internet of image things-discrete wavelet transform and Gabor wavelet transform based image enhancement resolution technique for IoT satellite applications," *Cognitive Systems Research*, vol. 57, pp. 46–53, 2019.
[11] T. Laksmi, T. Madhu, K. Kavya, and S. E. Basha, "Novel image enhancement technique using CLAHE and wavelet transforms," *International Journal of Scientific Engineering and Technology*, vol. 5, no. 11, pp. 507–511, 2016.
[12] A. Kaur, "A review on image enhancement with deep learning approach," *ACCENTS Transactions on Image Processing and Computer Vision*, vol. 4, no. 11, 16, 2018.
[13] Z. Chen, K. Pawar, M. Ekanayake, C. Pain, S. Zhong, and G. F. Egan, "Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges," *Journal of Digital Imaging*, vol. 36, no. 1, pp. 204–230, 2023.
[14] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *IEEE Transactions on Image Processing*, vol. 29, pp. 9140–9151, 2020.
[15] C. Wan, X. Zhou, Q. You, J. Sun, J. Shen, S. Zhu, Q. Jiang, and W. Yang, "Retinal image enhancement using cycle-constraint adversarial network," *Frontiers in Medicine*, vol. 8, 793726, 2022.
[16] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightenNet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognition Letters*, vol. 104, pp. 15–22, 2018.
[17] X. Kuang, X. Sui, Y. Liu, Q. Chen, and G. Gu, "Single infrared image enhancement using a deep convolutional neural network," *Neurocomputing*, vol. 332, pp. 119–128, 2019.
[18] K. Singh, and A. S. Parihar, "Dse-net: Deep simultaneous estimation network for low-light image enhancement," *Journal of Visual Communication and Image Representation*, vol. 91, 103780, 2023.
[19] P. M. Bentley, and J. McDonnell, "Wavelet transforms: an introduction," *Electronics & Communication Engineering Journal*, vol. 6, no. 4, pp. 175–186, 1994.
[20] N. Jin, and D. Liu, "Wavelet basis function neural networks for sequential learning," *IEEE Transactions on Neural Networks*, vol. 19, no. 3, pp. 523–528, 2008.
[21] C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang, "Multi-stage image denoising with the wavelet transform," *Pattern Recognition*, vol. 134, pp. 109050, 2023.
[22] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2021.
[23] P. Koranga, G. Singh, D. Verma, S. Chaube, A. Kumar, and S. Pant, "Image denoising based on wavelet transform using Visu thresholding technique," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 3, no. 4, 444, 2018.
[24] B. Xie, Z. Xiong, Z. Wang, L. Zhang, D. Zhang, and F. Li, "Gamma spectrum denoising method based on improved wavelet threshold," *Nuclear Engineering and Technology*, vol. 52, no. 8, pp. 1771–1776, 2020.
[25] T. Wiatowski, and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1845–1866, 2017.

[26] R. Koprowski, and S. Wilczyński, "Thermal image analysis using the serpentine method," *Infrared Physics & Technology*, vol. 89, pp. 97–109, 2018.
[27] F. Kong, J. Tang, Y. Li, D. Li, and K. Hu, "Dual-branch spectral–spatial feature extraction network for multispectral image compression," *Multimedia Systems*, vol. 29, no. 6, pp. 3579–3597, 2023.
[28] W. Kim, A.-D. Nguyen, J. Kim, J. Kim, H. Oh, and S. Lee, "Diverse and adjustable versatile image enhancer," *IEEE Access*, vol. 9, pp. 80883–80896, 2021.