Attention Mechanism-Enhanced Model for Automated Simple Brush Stroke Painting

Jiayin Zhang*
School of Culture and Tourism, Luoyang Polytechnic, Luoyang 471000, China
E-mail: zjy201953018@163.com
*Corresponding author

Keywords: attention mechanism, automation, hand drawn, simple brush painting techniques

Received: November 18, 2024

With the continuous development of computer technology, the application scope has become increasingly widespread. The flexible application of computer technology related mechanisms in automated simple brush painting technology can be further developed. Therefore, to further optimize the accuracy of automated simple stroke painting technology in generating images, making the images closer to real hand drawn images, a simple stroke painting model based on attention mechanism and long short-term memory network is constructed. The experimental environment is conducted using NVIDIA Tesla K80 GPU, 256GB of memory, and running on Python 3.8.13 and TensorFlow 2. X. On the FaceX dataset, the F1 score and Precision of the AM-LSTM reached 98.75% and 98.63% respectively. Compared to CNN, the F1 score and Precision has increased by 2.01% and 1.40% respectively. The improved attention mechanism model has good performance in automated simple brush painting technology. This indicates that the improved attention machine model can effectively improve image recognition performance. This research method innovatively combines attention mechanisms with long short-term memory network to construct the simple brush painting technology, resulting in more vivid and realistic simple strokes.

Povzetek: Predstavljen je model za avtomatizirano preprosto slikanje z uporabo mehanizma pozornosti in omrežja dolgoročne kratkoročne pomnilnosti (LSTM), ki izboljša prepoznavanje slik ter kvaliteto generiranja slik v primerjavi s tradicionalnimi metodami.

1 Introduction

After entering the computer age, Artificial Intelligence (AI) technology has a significant positive impact on various industries that cannot be ignored. The continuous development of various industries has also led to the widespread application of AI technology [1]. Simple strokes are abstract expressions of real objects. Simple lines are used to represent the contour features of real objects. Simple brush stroke painting plays an important role in daily life. The aim is to optimize the accuracy of automatic simple brush stroke painting technology in generating images, and truly achieve more accurate images generated by simple brush stroke painting technology [2]. Deep learning techniques play an important role in extracting line features from simple paintings. However, currently applying deep learning techniques to simple brush stroke painting generation has problems such as low efficiency and insufficient accuracy, as overly simple model structures may not fully capture stroke data features, leading to underfitting issues, while complex models can lead to overfitting [3]. Attention mechanism utilizes visual information to process resources. Non-attention mechanism models typically extract features through local operations, making it difficult to capture global contextual information. Meanwhile, it may perform poorly in handling complex feature relationships, as there may be complex spatial relationships between different regions. These relationships are difficult to effectively capture in local operations. Therefore, the application of attention mechanism in simple paintings can make it more convenient to obtain regional image features, thereby better capturing global and relatively complex feature relationships [4-5]. In view of this, this research has conducted in-depth analysis and exploration of attention mechanisms. It is applied to practical automated simple brush painting technology, aiming to make it sustainable.

2 Related works

Attention mechanism is a core technology applied in fields such as natural language processing, statistical learning, image detection, and speech recognition. Combining attention mechanisms with various neural networks can effectively achieve the processing and allocation of information resources. Hou Y et al. combined attention mechanisms with residual learning networks to identify continuous information between slices in medical image sequences. Its classification performance was significantly higher than that of 2D Convolutional Neural Network (CNN) and 3D CNN without attention mechanism. The average accuracy of this model was 88.69%, the average sensitivity was 87.58%, and the average specificity was 90.26% [6]. Yu et al. proposed a recursive neural network method based on volume aware positional attention (VPA-RNN) to capture data information in stock market trends. By adding

positional awareness to the attention mechanism, transaction volume was incorporated into the attention distribution mechanism. The experimental results showed that the proposed VPA-RNN could significantly outperform several existing highly competitive methods [7]. Zhu X et al. used attention mechanism to achieve image planning generation. The generator was integrated with conditional convolutions constrained by boundary inputs and attention modules with channel and spatial features. The experimental results showed that the method had high accuracy, demonstrating the effectiveness and superiority [8]. Xiao et al. proposed a neural information retrieval model that integrated attention mechanisms. A unified data fusion method was constructed based on the natural graphical features of the distribution system. The results indicated that this method had high accuracy in similarity matching of historical operating features and effectively supported intelligent fault diagnosis and troubleshooting of distribution systems [9]. To extract more discriminative features for hyperspectral images and prevent network degradation due to deepening, Xu et al. proposed a multi-scale spectral spatial fusion attention module for hyperspectral image classification. The channel optimization module was used to quantify the importance of feature maps at the channel level. Compared with other state-of-the-art deep networks, this method had higher overall accuracy (OA), average accuracy (AA), and Kappa coefficients [10].

With the development of AI technology and the diversification of expression needs, AI painting has rapidly developed. It has been widely applied in fields such as digital art, virtual reality, film and television production, and game development. Xu et al. proposed a structural variance model to analyze user behavior intention under cognitive conditions of AI painting. Firstly, AI painting was used to analyze user behavior data, classify psychological states, and remove irrelevant user intention data. Then, based on the classification results of user behavior data, it was compared with previous user willingness analysis methods. Compared with traditional methods, the proposed method had an accuracy of 86.5%, which shortened the prediction time of user behavior [11]. Zhou et al. proposed a two-stage coverage planning framework to automatically generate the optimal moving base path and robotic arm trajectory to draw walls. In the proposed framework, the global planner planed the sequence of painting route points. The local planner generated the mobile base pose through a new evaluation function. The on-site test results indicated that the entire coating robot system had high environmental adaptability, with better precision, recall and F1 value [12]. Zhang et al. proposed a new robust multi-view fuzzy clustering algorithm for image segmentation of Chinese literati paintings. This method was used to effectively decompose and extract ancient paintings. By effectively decomposing and extracting literati paintings, the electronic and digital transformation and preservation of literati paintings were achieved. Experiments showed that this algorithm could effectively segment and extract painting works, with lower errors [13]. Zhou et al. designed an automatic spraying monitoring and remote operation system based on digital twin technology. It could achieve real-time monitoring and remote fault handling during the spraying process. On this basis, a three-dimensional model of the environment and equipment was established. Based on the fusion and mapping of sensor data and geometric models, an augmented reality spray dual model was obtained. The results indicated that the system was more efficient than manual spraying methods [14].

The summary of relevant literature is shown in Table 1.

Table 1: Summary of relevant literature

Literature	Method	Results	Evaluation indicators,
Hou Y et al. [6]	Combined attention mechanisms with residual learning networks to identify continuous information	It performs better than 2D and 3D convolutional neural networks	Accuracy, sensitivity and specificity
Yu et al. [7]	A recursive neural network method based on volume aware positional attention to capture data information in stock market trends	It outperforms several existing highly competitive methods	Accuracy and error rate
Zhu X et al. [8]	An image planning generation based on attention mechanism	It has the best predictive performance	Accuracy
Xiao et al. [9]	a neural information retrieval model that integrates attention mechanisms.	It effectively supports intelligent fault diagnosis	Accuracy
Xu et al. [10]	A multi-scale spectral spatial fusion attention module for hyperspectral image classification	It performs better	Overall accuracy, average accuracy, and Kappa coefficients
Xu et al. [11]	A structural variance model to analyze user behavior intention	has an accuracy of 86.5%, which shortens the prediction time	Accuracy, prediction time
Zhou et al. [12]	A two-stage coverage planning framework to automatically generate the optimal moving base path and robotic arm trajectory	It has high environmental adaptability	Precision, recall and F1 value
Zhang et al. [13]	A new robust multi view fuzzy clustering algorithm for image segmentation of Chinese literati paintings.	It can effectively segment and extract painting works	Error
Zhou et al. [14]	An automatic spraying monitoring and remote operation system	The system is more efficient than manual spraying methods	Efficiency

In summary, attention mechanism plays a good role in image information processing. However, based on existing research, the focus is on automatic spraying, and the implementation technology of AI painting is relatively insufficient. Therefore, studying the advantages of attention mechanism in image information extraction and constructing automated simple drawing technology based on improved attention mechanism can better achieve the expression form of painting art and improve automated painting technology.

3 Construction of an automated simple brush stroke painting method based on improved attention mechanism

3.1 The basic mechanism of attention mechanism

To utilize visual information to process various resources, the human brain selects key areas in the visual area and focuses on them, which is the principle of attention mechanism [15]. The basic model of attention mechanism is shown in Figure 1.

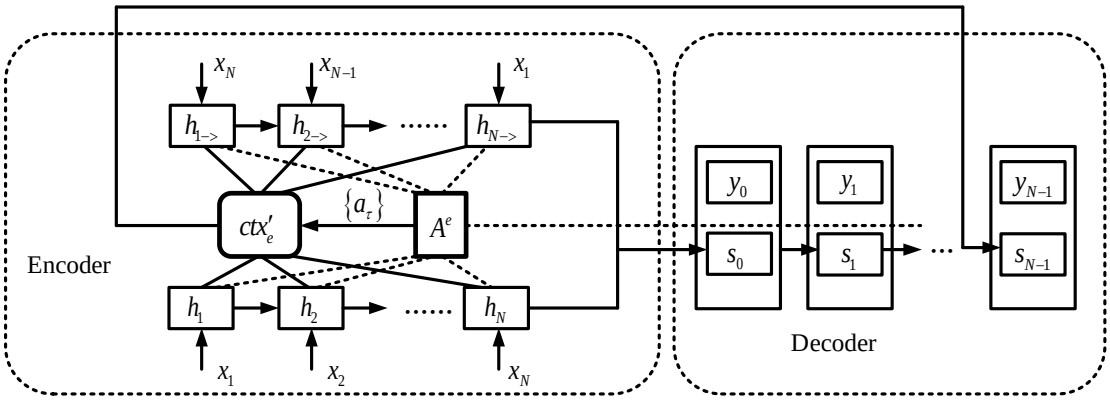


Figure 1: Attention mechanism model

In Figure 1, x represents the input data. y represents the output data. h represents the hidden state in the decoder. s indicates the hidden state in the decoder. According to the generation method, attention can be divided into two types. One is focused attention, which has active awareness and established goals, and focuses on a specific object. The second is attention based on saliency, which is not related to the target task of attention [16]. Attention can be divided into flexible attention and rigid attention in the own form. Flexible attention is an organic combination of attention values of different sizes, exhibiting their weight information in corresponding feature dimensions. Rigid attention is discrete positional information that is not combined. It only focuses on a specific location or dimension related to the corresponding input feature. Therefore, flexible attention has stronger performance and wider applications. The structure of the decoding model based on attention mechanism is shown in Figure 2.

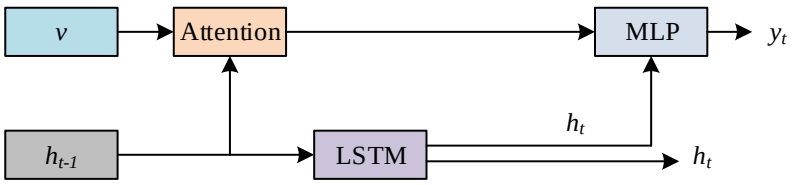


Figure 2: Decoding model based on adaptive attention mechanism

In Figure 2, V represents the set of image feature vectors in the model, and $V=\{v_1,v_2,\cdots,v_L\}$. At time t, the attention received by any region in the image depends on the state h_{t-1} of the hidden layer at time $t-1$. There are some differences between this model and the model based on traditional attention mechanisms. At this point, the weight of the attention mechanism is shown in equation (1) and equation (2).

$$z_t = w_h^T \tanh\left(W_v V + \left(W_g h_t\right)1^T\right) \tag{1}$$

$$\alpha_t = soft\max\left(z_t\right) \tag{2}$$

In equation (1), $W_v \in R^{L\times d}$, $W_g \in R^{L\times d}$, and $w_h \in R^L$ are three model parameters that need to be learned. $1\in R^L$ refers to a vector, with value of 1 for any element. h_t represents the hidden layer state of the Long

Short-Term Memory (LSTM) model in t. z_t represents the attention scoring function. In equation (2), α_t represents the vector composed of L attention weights. This vector corresponds to the region feature vector in the L image. A visual context vector can be calculated, as shown in equation (3).

$$\hat{v}_t = \sum_{i=1}^{L} \alpha_{ti} v_i \tag{3}$$

In addition, the language decoding model based on adaptive attention mechanism also introduces the concept of visual sentry. Visual sentry refers to the new components extracted from the model. This is because the memory unit of the language decoder has the ability to store relevant visual and language information. The model based on visual sentry self-attention mechanism is shown in Figure 3.

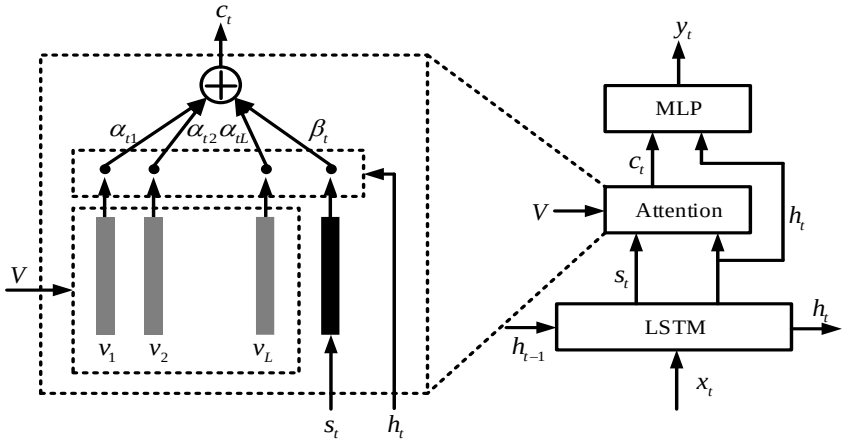


Figure 3: Self-attention mechanism model based on visual sentry

The role of the visual sentry is to accurately predict words as contextual information. In the task of image description generation based on attention mechanism, "visual sentry" is regarded as a component of the model. Through the attention mechanism, the model can dynamically decide whether to pay attention to new image regions in the process of image description generation, so as to improve the accuracy and naturalness of image description. The calculation equation for the vector s_t is shown in equations (4) and (5).

$$g_t = \sigma\left(W_x x_t + W_h h_{t-1}\right) \tag{4}$$

$$s_t = g_t \square \tanh\left(m_t\right) \tag{5}$$

In equation (4), σ represents the sigmoid function. W_x and W_g are the same, both representing the model parameters that need to be learned. Equation (4) plays an auxiliary role in calculating the equation (5). In equation (5), m_t represents the memory unit in the decoder. g_t represents the gate coefficient acting on m_t, which is used to process the relevant information in m_t. Then, the visual sentry s_t is obtained. Subsequently, based on the adaptive features of this mechanism, visual information and contextual information are screened. The weight distribution of regional feature vectors and visual sentry vectors can be calculated using equation (6).

$$\alpha_t = soft\max\left(\left[z_t; w_h^T \tanh\left(W_s s_t + W_g h_t\right)\right]\right) \tag{6}$$

In equation (6), $\alpha_t \in R^{L+1}$. The model parameters that need to be learned are W_s and W_g. Based on linear weighting, another context vector c_t can be calculated to summarize all input information. The calculation method is shown in equation (7).

$$c_t = \beta_t s_t + \left(1-\beta_t\right)\hat{v}_t \tag{7}$$

In equation (7), β_t represents a coefficient located within the $[0,1]$ interval. $\beta_t = \alpha_t\left[L+1\right]$. If the value of β_t is 0, the model needs to obtain relevant image information to obtain it. This reflects adaptive features that can intelligently select visual and contextual information. If the value of β_t is 1, the meaning is exactly the opposite. There is no need to obtain image information. This model only relies on language models, with a focus on visual sentry s_t.

3.2 A Sequence generation model for stroke brush painting process based on improved attention mechanism

In the field of computer vision, the research focuses on constructing models for visual attention mechanisms. The traditional feature integration theory has been extensively applied in this research, which has played a

powerful role in selecting concentrated visual features that cannot be ignored. The fused neural network structure can filter out the parts of the image region that have the most regional features. The key to traditional visual attention lies in obtaining the selection positions and methods in the relevant image regions. There is a strong correlation with the attention target. With the continuous development of deep learning, the application of deep neural networks is also becoming more widespread. This research organically combines the attention mechanism with it. The rigid attention described in section 2.1 is used to cut and obtain key areas of the image. Subsequently, the images obtained from this region are used to obtain the rigid attention position in the next stage, which is repeated several times. Rigid attention captures the key regions of an image, and first clarifies the specific area tasks for image processing. This helps identify the key areas that need to be extracted. Then, a pre-trained LSTM network is used to extract image features. Features that are relatively stable and not easily affected by factors such as lighting are selected from the extracted features, which are related to the rigid body. The importance of each feature for the target task is calculated. The attention weights are used to weight and fuse the extracted features. This can be achieved by multiplying the feature vectors of each feature point or region with their corresponding weights and adding the results. Rigid attention is beneficial for extracting task areas. Due to its feature stability, in image or video feature extraction, features have stronger consistency and recognizability under different perspectives, lighting, and poses. This helps the model to more accurately locate the task area, improving the accuracy and efficiency of feature extraction. After filtering all stages, the final image classification result can be obtained [17]. The attention mechanism can generate multiple attentions in the corresponding network and make them act on the relevant features separately. In the natural language processing framework, equation (8) can be obtained.

$$attention(Q,K,J) = soft\max\left(\frac{QK^T}{\sqrt{d_K}}\right)J \tag{8}$$

In equation (8), Q represents task related queries. K and J represent key value pairs for input features. The feature dimension corresponding to K is represented by d_K . Multi-terminal attention generates multiple attention through parallel action. All newly generated attention effects are jointly processed. The final output can be obtained, as shown in equation (9).

$$MultiHead(Q,K,J) = Concat(head_1,\ldots,head_h)W^O \tag{9}$$

In equation (9), $head_i$ corresponds to the result of any single attention action, as shown in equation (10).

$$head_i = attention\left(QW_i^Q, KW_i^K, VW_i^J\right) \tag{10}$$

According to the analysis, W_i^Q , W_i^K , and W_i^J in equation (10) are all projection matrices of image features. This model proves that the attention is usually only a single channel feature representation in most models based on flexible attention. Therefore, multi-channel and multi-dimensional attention information is constructed to enhance the relevant performance of attention mechanisms. The attention mechanism is integrated with the original sequence generation model, which can more accurately assist with simple brush stroke painting. The generation model is shown in Figure 4.

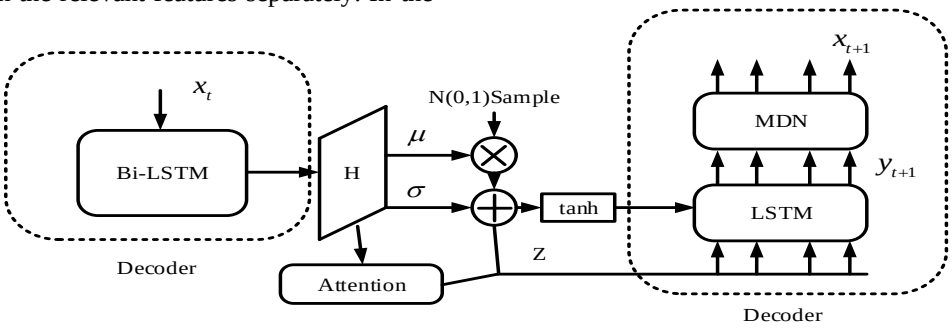


Figure 4: A simple brush stroke painting generation model based on attention mechanism and original sequence

In Figure 4, tanh represents the activation function. x_t represents input data. x_{t+1} represents the input data at time $t+1$. μ , and σ represent the weights to be optimized. Z is the intermediate variable. y_{t+1} represents the output data at time $t+1$. H represents the attention mechanism. According to Figure 4, combining attention mechanism with the generation model can successfully make the intermediate variable perform an attention mechanism calculation first. In the attention mechanism, intermediate variables connect input, output and the calculation of attention weight. The research uses query, key and value as intermediate variables for calculation. A query represents the part of information that currently needs attention or processing. It can be the output from the upper layer or the specific representation of the current task. The key is used to match the query to determine which input information is most important for the current task. The value is the actual information content associated with the key. Intermediate variables enable the model to dynamically focus on the key

information in the input data, thereby improving the performance of the model. Therefore, it is necessary to use intermediate variables to perform attention mechanism calculation. Subsequently, it is combined with the input of the image decoder LSTM. The pseudocode for AM-LSTM is displayed in Figure 5.

```
Initialize parameters
Input_dim hidden_dim, attention_dim ,image_size = (height, width)
Initialize weights
lstm_weights = initialize_weights
attention_weights = initialize_weights
Initialize LSTM state and attention context vector
lstm_state = initialize_lstm_state(hidden_dim)
attention_context = initialize_attention_context(attention_dim, image_size)
For y in range(height):
For x in range(width):
Attention_weights_current = compute_attention_weights(lstm_state,
attention_context, attention_weights, image_size, (y, x))
Context_vector = compute_context_vector(attention_weights_current,
attention_context)
Concatenate(lstm_state[-1], context_vector)
LSTM_state[-1], the last hidden state
LSTM_step(lstm_input, lstm_state, lstm_weights)
Pixel_value = generate_pixel_value(lstm_state[-1])
Attention_context = update_attention_context(attention_context, (y, x),
pixel_value)
Generated_image[(y, x)] = pixel_value
```

Figure 5: The pseudocode for AM-LSTM

The final model can provide sequences at any different time points based on prior knowledge, thereby generating targets with different weights. The generation of this model mainly relies on the basic framework of combining encoder and decoder. The Encoder-Decoder framework consists of two parts: an encoder and a decoder. The encoder is responsible for converting input sequences (such as images) into intermediate state vectors. The decoder gradually generates an output sequence for the current time state (such as a simple stroke) based on this intermediate state vector and the output of the previous time state. This framework can be extended based on specific tasks. In the generation of simple brush strokes, the encoder-decoder framework completes the generation task based on the specific content of the simplified drawing. The encoder converts the input simple stroke related data into feature vectors. The decoder generates the feature vectors of the generated simplified drawing data to generate simplified drawings. The combination of the two transforms relatively complex data samples into specific feature vectors and generates new simplified drawings based on these features. During the encoding process, CNN is usually used to effectively extract regional image features. Through relevant pre training, image features with strong generalization ability can be obtained. Subsequently, it is applied to solve various types of computer vision related problems [18]. The schematic diagram of the image encoder extracting regional image features is shown in Figure 6.

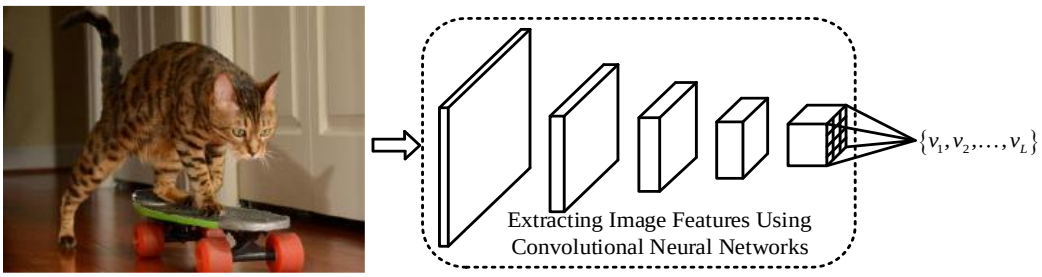


Figure 6: Schematic diagram of image encoder

Figure 6 illustrates the process of extracting regional image features by the network. The feature map output by the last convolutional layer in the network can represent the significant features of the target image. Subsequently, it is subjected to adaptive pooling processing to unify the size of all target images. The size of the image feature map is uniformly set to $14\times14\times2048$, taking into account various factors. A smaller feature map size (14×14) can reduce computational complexity and memory consumption, which can reduce parameter count and improve computational speed. Setting the depth of the feature map to 2048 channels ensures that the model captures sufficient image feature information without losing important details due to its small size. Therefore, there is also a certain value for the size of the image feature matrix $V=\{v_1,v_2,\cdots,v_L\}$. The correlation value of the feature vector corresponding to any image region is also determined, that is, the dimension v_i of is $d=2048$. All image regions support pre-extraction and permanent preservation of features. After the encoding is completed, another data containing prior information related to the original process sequence, namely hidden variables, can also be obtained. Introducing hidden variables into the decoder can effectively capture key information in the input sequence and pass it on to the decoder. In addition, on the basis of the hidden variable, the decoder can perceive contextual information in the input sequence, ensuring that each output generated by the decoder can take into account the entire input sequence information. Hidden variables are often used as the initial state variables of image decoders. It can also be introduced to the image decoder at any time. The process of introducing hidden variables into the decoder is as follows. The encoder first processes the input sequence and converts it into one or more hidden states. The hidden state of the encoder is transmitted directly to the decoder. The decoder calculates attention weights based on the hidden state of the current time step and the hidden state of the encoder, dynamically selecting the information with the highest correlation. After receiving the hidden state of the encoder, the decoder generates the current output word based on these states. Therefore, it can ignore the negative impact caused by the increase in time series.

The attention mechanism is fundamental in the entire generation model of sketch sequences. It is related to the accuracy and efficiency of the entire generation model. The hidden state synthesis of the decoder is represented as h_i. The hidden state at the previous moment is represented by s_{j-1}. Variation Autoencoders (VAE) can learn latent representations of image or text data by training VAE models and generate new data by sampling from the latent space, which is of great significance for fields such as data augmentation and artistic creation. Meanwhile, VAE has feature learning and dimensionality reduction capabilities. High dimensional data is compressed into a low dimensional space while preserving the main features of the data. The loss function of the generative model based on VAE includes KL divergence and reconstruction error. KL divergence is used to measure the distance between the probability distribution of the hidden variables encoded by the encoder and the true distribution. Reconstruction error can measure the error between the real sequence and the generated sequence. For the optimized attention mechanism model, to achieve the process sequence generation of simple brush painting, a portion of KL divergence is extracted from the target network model to avoid sequences that do not conform to the same Gaussian distribution converging into the same distribution sequence. Therefore, to measure the sequence generation model used in the study, it is necessary to comprehensively consider KL divergence and reconstruction error.

4 Effect analysis of automated simple brush painting method based on improved attention mechanism

4.1 Performance verification of automated simple brush painting method based on improved attention mechanism

To verify the performance of the simple brush painting sequence generation model based on attention mechanism, relevant training experiments are conducted in the research. The parameter settings for the AM-LSTM model are as follows. The learning rate in attention mechanism is set to 0.001, the Batch Size is 64, and the optimizer is Adam. The Hidden Units in the LSTM network are 128, the LSTM layers are 2, the learning rate is 0.001, and the Batch Size is 64. The epoch is 40,000, the final reconstruction error is taken as the loss function of the new algorithm, and the optimizer is Adam. The number of training times required to achieve the optimal solution of the loss function can be obtained. Furthermore, it provides reference for the subsequent generation of simple brush painting. The results are shown in Figure 7. The training loss value shows significant fluctuations between 0 to 15000 training cycles. Especially in the early training, the change amplitude is significant, constantly hovering between positive and negative values. After 15000 training sessions, the loss gradually decreases. The fluctuation of the loss function may be due to noise or imbalance in the training data, which affects the stability of the model. In addition, the model structure may also affect the volatility of the loss function. Complex model structures may be more susceptible to hyperparameters, leading to unstable training processes. The fluctuation is common in the training process of machine learning models. The loss value tends to flatten out. Then, there are no significant fluctuations. Finally, after 40000 training sessions, the trend curve of the loss value for the loss function is parallel to the horizontal axis, indicating that the loss function has reached the optimal solution.

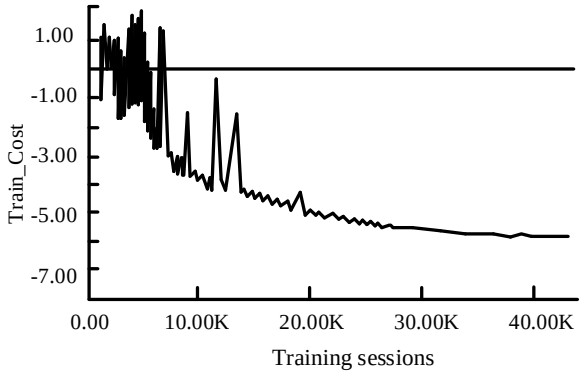


Figure 7: Trend of loss value changes in training experiments

Bilingual Evaluation Understudy (BLEU) Bleu-1, Bleu-3, METROR, CIDEr and other indicators are used to evaluate whether the generated simple strokes are related to the image. BLEU considers brushstrokes, colors, or composition elements as "vocabulary", so BLEU scores may reflect the degree to which these elements are reproduced in the work and the similarity to standard works. METEOR can evaluate the similarity of stroke style, color matching, or composition layout. CIDEr can calculate the similarity between stroke features (such as line thickness, direction, density, etc.) in painting works and standard works. The above indicators can better capture the uniqueness and information content in painting, as they consider the weight and distribution of stroke features. Figure 8 shows the variation curves of Bleu-1, Bleu-3, METEOR, CIDEr, etc. with iteration period. From Figure 8, each indicator shows a fluctuating upward trend. After 40 epochs, the indicators show significant fluctuations and gradually stabilize. The method proposed in the research performs well in various evaluation indicators. The obtained simplified stroke image has a high degree of fit with the actual image. This is because the proposed method introduces hidden variables in the attention mechanism to better capture the feature information of sample data, resulting in a better fit between the generated image and the real image. The performance is effectively improved.

To further verify the impact of attention mechanism on image feature extraction performance, the effectiveness of the improved attention mechanism method is demonstrated through ablation experiments. The method proposed in the study is represented as AM-LSTM. The dataset used for the test is from the Intelligent Big Data Visualization Laboratory (iDVX Lab) of Tongji University, a high-quality simple stroke dataset containing over 5 million cartoon facial expressions - FaceX [19]. This dataset is drawn and generated by a professional designer. To ensure data quality, preprocessing is first carried out, including missing sample data and abnormal data. Based on the characteristics and distribution patterns of the data, the noisy data is removed and the data purity is enhanced.

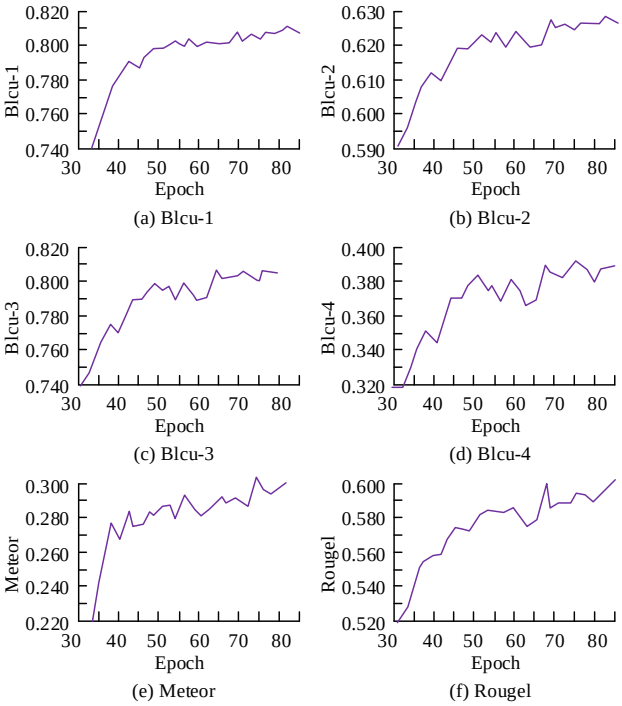
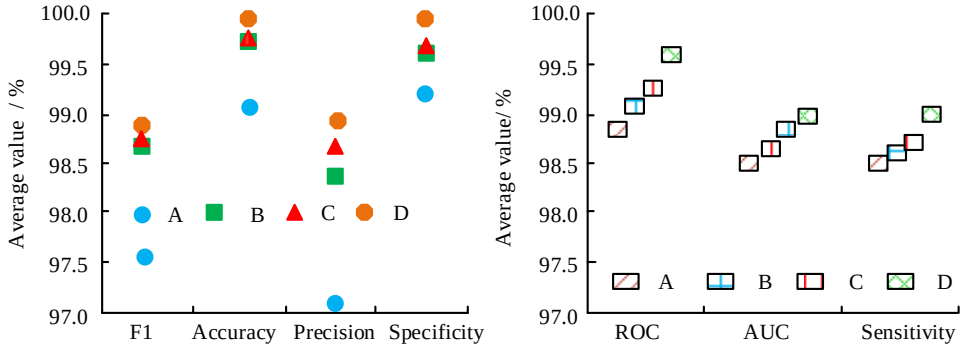


Figure 8: Performance comparison under different evaluation indicators

For missing data, relevant data is collected again to fill in the gaps. Then, the average value is taken to replace the outlier data. Next, the data is normalized to eliminate the influence of variable dimensions. This dataset is all in SVG format, fully recording every stroke during the drawing process. The SVG data format has a certain impact on model performance. SVG files are smaller than traditional image formats such as PNG or JPEG, which can reduce model loading time and resource consumption. In addition, SVG icons are based on XML code, which means that the color, shape, size, and other attributes of the icons can be directly modified through code, making the model more flexible and adaptable. The basic attention mechanism test is recorded as Experiment A. The CNN is denoted as Experiment B. The combination of multi-scale feature extraction and AM network is recorded as Experiment C. The proposed AM-LSTM model is recorded as Experiment D. The basic attention mechanism test refers to using only attention mechanism to generate images. The image description generation is carried out using the "encoding-decoding" method. Multi-scale feature extraction uses different scales (or resolutions) to capture image features, including edges, textures, shapes, etc. By converting these features into high-dimensional data, they can be further used for tasks such as image analysis, object detection, and image classification. The comparison of the four experimental results is shown in Figure 9. Figure 9 (a) shows the F1, accuracy, precision, and specificity results of four experiments. Figure 9 (b) shows the ROC, AUC, and sensitivity results of four experiments. From Figure 9 (a), the F1, accuracy, precision, and specificity values of Experiment D are 97.85%, 98.74%, 97.74%, and 98.77%, respectively.

The improved attention mechanism model is superior to the other three models. In Figure 9 (b), the ROC, AUC, and sensitivity indicators of Experiment D are better than the other three experiments, indicating that improving the attention machine model can effectively improve image recognition performance. From Figure 9, in the comparison experiment, the F1, accuracy, precision, and specificity indicators all perform better.



(a) F1, Accuracy, Precision and Specificity results

(b) ROC, AUC and Sensitivity results

Figure 9: Mean value of performance evaluation indexes of four experiments

To further validate the performance of the proposed simple stroke image processing model, commonly used image recognition models are selected for comparison, including CNN, Recurrent Neural Networks (RNN), and Random Forest (RF). The performance comparison results of the four methods in this dataset are shown in Figure 10. From Figure 10, the average performance evaluation indicators of the AM-LSTM proposed in the study are superior to the other three methods. The F1 score and precision reach 98.75% and 98.63% respectively. Compared to CNN, it has increased by 2.01% and 1.40% respectively. Meanwhile, the accuracy of the proposed AM-LSTM model has increased by 0.51%, 0.38%, and 0.24% compared to CNN, RNN, and RF, respectively. The precision is 1.68% and 1.53% higher than RNN and RF, respectively. Overall, the performance improvement of AM-LSTM is more significant.

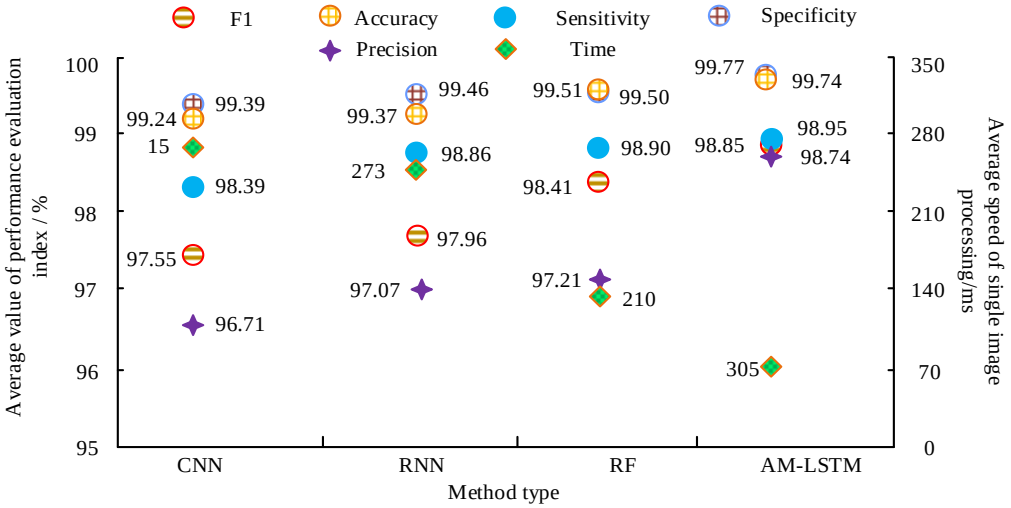


Figure 10: Performance evaluation index mean of four methods in FaceX database

4.2 The application effect analysis of automated simple brush painting method based on improved attention mechanism

The sample generated based on the proposed method is shown in Figure 11. When using a native VAE framework for simple brush painting, models based on attention mechanisms are easier to generate simple brush painting works that meet the requirements. This simple brush painting work is more complete and accurate.

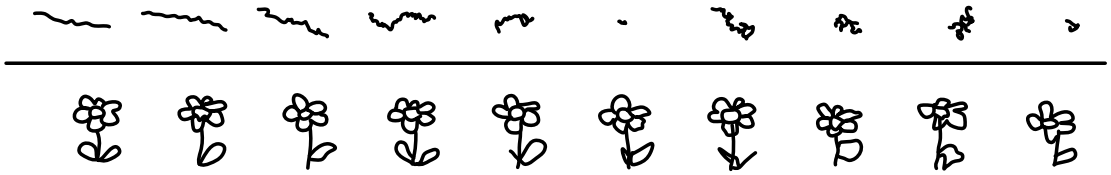


Figure 11: Comparison of image generation effects between the original model and the experimental model

However, it should be noted that there are still some shortcomings in the attention mechanism based simple brush painting algorithm. There is a lack of a unified evaluation and judgment standard for the generated simple brush painting works. Based on this, the Turing test is to conduct a detailed effect evaluation on the generated painting works. When conducting the Turing test, the first step is to choose a suitable text conversation platform, namely, the instant messaging software. Then, the required data samples are prepared for testing, which are related works and hand drawn works generated by network automated simple drawing technology based on attention mechanism. The false positive rate determines which sample data is a machine and which is a human hand drawn artwork. If the false positive rate reaches a certain percentage (over 30%), it can be considered that the machine has passed the Turing test. The results are shown in Figure 12. The first line in Figure 12 shows the simple brush paintings in the dataset. The third row is the finished product of the corresponding simple brush painting works in the dataset after processing. It refers to the related works generated by network automated simple brush painting technology based on attention mechanism. The relevant data in the second and fourth lines refer to the proportion of their corresponding simple brushstroke paintings that are considered real hand-painted works. By analyzing this data, besides the third simple brush painting work, the simple brush painting works generated by automated simple brush painting technology based on attention mechanism are more likely to be considered as hand drawn works. It has a certain degree of significance.

Dataset				
False positive rate	42.86%	12.93%	55.10%	26.53%
Attention network				
False positive rate	57.14%	87.07%	49.90%	73.47%

Figure 12: Turing test results

To further explore the generation accuracy of automated simple brush painting technology based on attention mechanism, Turing testing is conducted on simple brush painting works generated by multiple types of models. Moreover, the difficulty of drawing simple brush painting model images is greater. In Figure 13, A, B, and C represent different types of algorithms. A represents the corresponding model in the dataset. B is the algorithm model optimized for selecting regional features. C is the proposed algorithm model based on attention mechanism. If the false positive rate reaches a certain percentage (over 30%), it can be considered that the machine has passed the Turing test. The test results of different methods in the three paintings are shown in Figure 13. The second, fourth, and sixth lines are the Turing test results corresponding to the three types mentioned above. Based on a detailed analysis of the data in each row, the automated simple brush painting technology based on attention mechanism is superior. The generated simple brush painting works are more vivid and realistic. Its false positive rate exceeds 30% and it has passed the Turing test.

A			
False positive rate	5.2%	6.8%	10.3%
B			
False positive rate	1.3%	1.5%	1.5%
C			
False positive rate	42.6%	51.5%	39.6%

Figure 13: The Turing test results comparison for generating multiple types of models

To further standardize the evaluation of various benchmark methods, the FaceX dataset is taken as an example for further evaluation. The comparison methods include attention mechanism, Dual attention network (DANet), Efficient Channel Attention Module (ECA) and AM-LSTM. The evaluation metrics are Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR). The results are shown in Table 2. In the comparison of SSIM, attention mechanism, DANet, and ECA are 0.55, 0.76, and 0.49, respectively. The SSIM of GM-LSTM is 0.83. In the comparison of PSNR, attention mechanism, DANet, and ECA are 14.52, 18.39, and 18.08, respectively The PSNR of GM-LSTM is 22.91. Overall, the research method performs better on SSIM and PSNR, and the generated images have better quality.

To verify the performance of the AM-LSTM, the significance test is conducted. Benchmark methods attention mechanism and LSTM are introduced for comparison. The obtained *P*-values, χ^2 test and Confidence Interval (CI) results are shown in Table 3. The research method has stability in the test results of different indicators. Although the *P*-values for attention mechanism and LSTM are significant, the significance is low, and the model performance is significantly lower than that of the research method. After comprehensive verification, this research method performs better among various comparison methods, verifying its effectiveness in simple brush painting.

Table 2: Test results for SSIM and PSNR

Model	SSIM	PSNR
Attention mechanism	0.55	14.52
DANet	0.76	18.39
ECA	0.49	18.08
AM-LSTM	0.83	22.91

Table 3: Statistical validation results of the research model

Testing index	Research method			Attention mechanism			LSTM		
Testing index	*P*	χ^2	CI	*P*	χ^2	CI	*P*	χ^2	CI
Precision	0.001	5.277	0.94	0.01	4.851	0.86	0.02	5.614	0.92
Recall	0.002	8.419	0.85	0.04	9.882	0.89	0.01	3.092	0.91
F1	0.001	9.064	0.91	0.05	6.818	0.91	0.01	8.572	0.85

5 Discussion

This study combines attention mechanism and LSTM to construct an AM-LSTM model for generating simple brushstrokes. After multiple experimental verifications, the research method has shown good performance in multiple indicators. Specifically, in Bleu-1, Bleu-3, METEOR, CIDEr, the proposed method performs well in various evaluation indicators. The obtained simplified stroke image has a high degree of fit with the actual image. This is because the proposed method introduces hidden variables in the attention mechanism to better capture the feature information of sample data, resulting in a better fit between the generated image and the real image. In addition, The F1 score and precision of the AM-LSTM reach 98.75% and 98.63% respectively. Compared to CNN, it has increased by 2.01% and 1.40% respectively.

Meanwhile, the accuracy of the proposed AM-LSTM model has increased by 0.51%, 0.38%, and 0.24% compared to CNN, RNN, and RF, respectively. From this perspective, the research method performs better in various indicators. In the Turing test, the generated simple brush painting works are more vivid and realistic. Its false positive rate exceeds 30%, which indicates that the designed method passed the Turing test. In the SSIM and PSNR metric tests, the research methods are 0.83 and 22.91, respectively. This is because introducing hidden variables in the decoder can effectively capture key information in the input sequence and pass it to the decoder, ensuring that the decoder can consider the information of the entire input sequence relatively comprehensively. In the current study, Wang X et al. constructed a multi-level Generative Adversarial Network (GAN) architecture. It consists of three different GANs that independently model structural, semantic, and texture patterns to improve the fidelity and stability of the generation process [20]. The method proposed in this research has similar conclusions to it. In addition, Yan S et al. utilized attention mechanisms to extract long-distance and irregular image content, resulting in the generation of complete images. The experiment shows that the results generated by this method are more natural and realistic, and the completed parts exhibit more connection consistency [21]. Overall, utilizing attention mechanisms to construct image generation models has achieved some research results. Comparatively speaking, the research method has shown better performance in multiple indicators and can better achieve the generation of simple brushstrokes, with certain practical significance.

6 Conclusion

Modern science and technology have greatly promoted the intelligent development of various industries. Attention has higher intelligence and stronger ability to extract image information. To achieve good and sustainable development of automated simple brush painting technology, attention mechanism is applied. Then, the LSTM was used to extract sequence features. The results show that the average performance evaluation indicators of the AM-LSTM network proposed in the research are superior to the other methods. The F1 score and precision reached 98.75% and 98.63% respectively. Compared to CNN, it has increased by 2.01% and 1.40% respectively. The Turing test results show that the simple brush painting works generated by automated simple brush painting technology based on attention mechanism are more likely to be considered as hand drawn works. This method passed the Turing test. Overall, in the tested data samples, the proposed automated simple drawing technique can generate more realistic painting works. Although this study fortunately achieved some research results, there are still some shortcomings. Firstly, the number of Turing tests is relatively small. Secondly, the generalizability of the experimental results is limited by the sample data. In the future, these aspects can be optimized. The research method is further optimized to handle more complex art styles or real-time processing. The achievements are applied more widely. In addition, specific methods or experimental conditions that can be explored, such as expanding the diversity of the dataset or integrating other AI technologies, and generative adversarial networks for enhanced realism.

Reference

[1] Mathew L, V. R. B. Efficient Transformer Based Sentiment Classification Models. Informatica, 2022, 46(8). DOI: 10.31449/inf.46.8.1234

[2] Li Z, Qian Y, Wang H, Zhou X L, Sheng G H, Jiag X C.A novel image-orientation feature extraction method for partial discharges. IET Generation, Transmission and Distribution, 2022, 16(6): 1139-1150. DOI: 10.1049/iet-gtd.2021.1234

[3] Li P P, Zeng G H, Bo H, Ling Y, Shi Z C, He C P, Liu W, Chen Y. A short text classification model for electrical equipment defects based on contextual features. Wuhan University Journal of Natural Sciences, 2022, 27(6): 465-475. DOI: 10.1007/s11859-022-1234-5

[4] Zhang J, Zhou Y, Xia K, Jiang Y, Liu Y. A novel automatic image segmentation method for Chinese literati paintings using multi-view fuzzy clustering technology. Multimedia Systems, 2020, 26(1): 37-51. DOI: 10.1007/s00530-019-1234-6

[5] Muneer A, Dairabayev Z. Design and implementation of automatic painting mobile robot. IAES International Journal of Robotics and Automation (IJRA), 2021, 10(1): 68-74. DOI: 10.11591/ijra.v10i1.1234

[6] Hou Y, Su J, Liang J, Chen X, Liu Q, Deng L, Liao J. A stroke image recognition model based on 3D residual network and attention mechanism. Journal of Intelligent & Fuzzy Systems, 2022, 43(4): 5205-5214. DOI: 10.3233/JIFS-212345

[7] Yu X, Li D, Shen Y. Forecasting stock index using a volume-aware positional attention-based recurrent neural network. International Journal of Software Engineering and Knowledge Engineering, 2021, 31(11n12): 1783-1801. DOI: 10.1142/S0218194021123456

[8] Zhu X, Liu Y, Liang L, Wang T, Li Z, Deng Q, Liu Y. Multiple Layout Design Generation via a GAN-Based Method with Conditional Convolution and Attention: Regular Section. IEICE Transactions on Information and Systems, 2023, E106.D(9):1615-1619. DOI: 10.1587/transinf.2022EDP1234

[9] Xiao K, Li D, Guo P, Wang X H, Chen Y. Similarity matching method of power distribution system operating data based on neural information retrieval. Global Energy Interconnection, 2023, 6(1): 15-25. DOI: 10.1016/j.gloei.2023.01.002

[10] Xu Q, Liang Y, Wang D, Luo B. Hyperspectral image classification based on SE-Res2Net and multi-scale spatial spectral fusion attention mechanism. Journal of Computer-Aided Design and Computer Graphics, 2021, 33(11): 1726-1734. DOI: 10.3724/SP.J.1089.2021.1726

[11] Xu J, Yoo C, Pan Y. Prediction of user behavioral intentions based on structural equation modelling in AI painting cognitive conditions. International Journal of Communication Networks and Information Security, 2023, 15(1): 147-153. DOI: 10.1504/IJCNIS.2023.10012345

[12] Zhou Y, Li P, Ye Z F, Yue L Z, Gui L H, Jiang X, Li Xiang Liu Y H. Building information modeling-based 3D reconstruction and coverage planning enabled automatic painting of interior walls using a novel painting robot in construction. Journal of Field Robotics, 2022, 39(8): 1178-1204. DOI: 10.1002/rob.22045

[13] Zhang J, Zhou Y, Xia K, Jiang Y Z, Liu Y. A novel automatic image segmentation method for Chinese literati paintings using multi-view fuzzy clustering technology. Multimedia Systems, 2020, 26(1): 37-51. DOI: 10.1007/s00530-019-1234-6

[14] Zhou K, Yang S, Guo Z, Long X J, Hou J L, Jin T G. Design of automatic spray monitoring and tele-operation system based on digital twin technology. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2021, 235(24): 7709-7725. DOI: 10.1177/09544062211012345

[15] Liang Y. Analysis of the integration of Chinese painting techniques in watercolor painting. Arts Studies and Criticism, 2022, 3(1): 37-40. DOI: 10.1016/j.asc.2022.01.003

[16] Karusine A H, Cui C, Larsey N O, Nolain T K. The impact of automation in painting furniture parts. Open Journal of Applied Sciences, 2022, 12(12): 1995-2003. DOI: 10.4236/ojapps.2022.1212345

[17] Kesiman M W A, Dermawan K T. AKSALont: Automatic transliteration application for Balinese palm leaf manuscripts with LSTM Model. Jurnal Teknologi dan Sistem Komputer, 2021, 9(3): 142-149. DOI: 10.14710/jtsiskom.9.3.142-149

[18] Yan X Y. Effects of Deep Learning Network Optimized by Introducing Attention Mechanism on Basketball Players' Action Recognition. Informatic, 2024, 48(19). DOI: 10.1016/j.informatic.2024.123456

[19] Rinki K, Verma P, Choudhury T, Singh B K. A novel feature extraction dual DCT-DWT image watermarking combined with chaos-based cryptosystem. Journal of Computer Sciences, 2021, 17(10): 971-983. DOI: 10.3844/jcssp.2021.971-983

[20] Wang X, Hui B, Guo P, Jin R, Ding L. Coarse-to-Fine Structure and Semantic Learning for Single-Sample SAR Image Generation. Remote Sensing, 2024, 16(17): 3326-3326. DOI: 10.3390/rs16173326

[21] Yan S, Zhang X F. PCNet: partial convolution attention mechanism for image inpainting. International Journal of Computers and Applications, 2022, 44(8): 738-745. DOI: 10.1080/1206212X.2022.1234567