



OPEN

## The artistic image processing for visual healing in smart city

Guangfu Qu<sup>1</sup>✉, Qian Song<sup>2</sup> & Ting Fang<sup>2</sup>✉

This study investigates the processing methods of artistic images within the context of Smart city (SC) initiatives, focusing on the visual healing effects of artistic image processing to enhance urban residents' mental health and quality of life. Firstly, it examines the role of artistic image processing techniques in visual healing. Secondly, deep learning technology is introduced and improved, proposing the overlapping segmentation vision transformer (OSViT) for image blocks, and further integrating the bidirectional long short-term memory (BiLSTM) algorithm. An innovative artistic image processing and classification recognition model based on OSViT-BiLSTM is then constructed. Finally, the visual healing effect of the processed art images in different scenes is analyzed. The results demonstrate that the proposed model achieves a classification recognition accuracy of 92.9% for art images, which is at least 6.9% higher than that of other existing model algorithms. Additionally, over 90% of users report satisfaction with the visual healing effects of the artistic images. Therefore, it is found that the proposed model can accurately identify artistic images, enhance their beauty and artistry, and improve the visual healing effect. This study provides an experimental reference for incorporating visual healing into SC initiatives.

**Keywords** Smart city, Deep learning, Visual healing, Artistic image, Visual attention model

### Research background and motivations

Although the rapid development of Smart city (SC) has provided urban residents with abundant data for daily life and work, the pressures from employment and living conditions continue to pose challenges to their mental health. For example, the accelerated urban pace has compounded these pressures, resulting in increased psychological stress among residents<sup>1–3</sup>. Therefore, there is a growing emphasis on mental health and social well-being in SC development. Consequently, exploring artistic image processing technology with visual healing effects in SC construction holds significant practical implications for residents' mental health.

Visual healing employs color, shape, and line from artistic images to improve people's mental health<sup>4</sup>. Research increasingly supports its efficacy in alleviating stress, fatigue, and anxiety among urban populations<sup>5–7</sup>. By processing and designing artistic images in urban environments, comfortable and pleasant urban spaces can be created that help relieve residents' anxiety and stress<sup>8</sup>. Moreover, integrating art image processing technology with visual healing in SC development can provide residents with personalized and intelligent visual healing services, thus enhancing urban residents' quality of life and well-being<sup>9,10</sup>. Hence, the art image processing technology of visual healing in SC has vital application necessity and development prospects.

### Research objectives

The main purpose of this study is to discuss the artistic image-processing technology of visual healing in SC and propose strategies and schemes based on the SC vision. The innovations of the study are multifaceted. First, the relevance of visual healing in SC is expounded by analyzing current urban residents' mental health challenges and existing issues. Second, the application of art image processing technology in visual healing is introduced, presenting a model for art image processing and classification recognition based on Overlapping Segmentation Vision Transformer (OSViT)-Bidirectional Long Short-Term Memory (BiLSTM). Lastly, aligned with the SC development vision, the study proposes an implementation scheme for visual healing in SC and assesses its feasibility. This study offers experimental reference direction for the research on residents' mental health within the SC vision.

<sup>1</sup>School of Film, Shandong University of Arts, Jinan 250300, China. <sup>2</sup>Sichuan Fine Arts Institute, Chongqing 400053, China. ✉email: qgf1974@163.com; xianouxi810@163.com

## Literature review

In recent years, with the development of SC and increased attention to health, mental health healing has gradually become a research hotspot. Freeman et al.<sup>11</sup> evaluated a meditation-based mental stress relief plan using the Center for Disease Control framework, offering mental healing for veterans' hospital employees. The findings suggested that the program prominently mitigated anxiety and depression levels while enhancing employee well-being and responsible behavior. Hass-Cohen et al.<sup>12</sup> discussed the efficacy of painting as a visual healing method for patients with chronic diseases. They believed that art therapy effectively reduced patients' pain, depression, and anxiety and improved the quality of interpersonal relationships. Bowen-salter et al.<sup>13</sup> applied art visual healing to patients with Post-Traumatic Stress Disorder (PTSD). They demonstrated that as a trauma treatment intervention, it could be paired with psychotherapy, including humanistic therapy and family therapy, to achieve good treatment outcomes. Lee et al.<sup>14</sup> designed a multimodal food art therapy program, discovering that it had therapeutic effects on individuals with mild cognitive impairment and mild dementia. Julianitino et al.<sup>15</sup> introduced the virtual healing environment developed on virtual reality platforms, focusing on virtual environment design and technical implementation to provide an immersive therapeutic experience, with potential applications in psychotherapy. Skorburg et al.<sup>16</sup> paid attention to the ethics of mental health research, discussing the ethical issues and participatory turn of artificial intelligence (AI) in this field. Mitro et al.<sup>17</sup> proposed an AI-based smart bracelet to monitor physiological indicators and stress levels in real-time.

Artistic image processing and classification are pivotal in visual healing, attracting significant attention from domestic and foreign scholars. Ayana et al.<sup>18</sup> proposed a novel multi-stage transfer learning method for classifying breast cancer ultrasound images, demonstrating superior accuracy and reliability compared to traditional methods. Maniat et al.<sup>19</sup> utilized a deep learning (DL)-based method to visualize crack detection using Google Street View images, achieving high detection accuracy. Wang et al.<sup>20</sup> proposed a microcapsule self-healing mortar object state recognition method based on an X-ray Computed Tomography image, effectively identifying microcapsule conditions. Batziou et al.<sup>21</sup> developed a neural style transfer method based on CycleGAN for art, enabling style transfer through adaptive information selection. Tehsin et al.<sup>22</sup> proposed a scalable DL-based method for satellite image classification, demonstrating robust classification performance across multiple satellite image datasets.

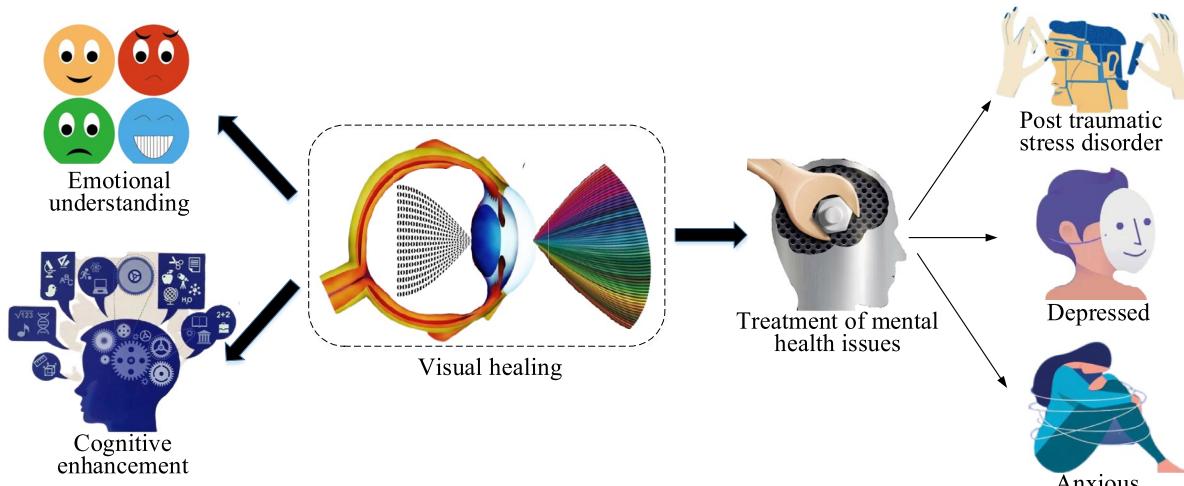
In summary, the analysis of previous studies reveals that while many studies have involved psychological healing through vision, the connection between visual healing and artistic image processing remains the current research gap. Meanwhile, in the domain of image processing, DL methods have predominantly been applied to medical and satellite image classification, such as Ayana et al.<sup>18</sup>, Wang et al.<sup>20</sup>, and Tehsin et al.<sup>22</sup>. However, there is a noticeable scarcity of research on art image processing. Therefore, this study addresses this gap by employing DL techniques for art image processing and evaluating their efficacy in enhancing the impact and treatment efficiency of visual healing within SC environments. This contribution aims to advance the health and well-being initiatives within SC.

## Research model

### Visual healing analysis

Visual healing is a psycho-emotional approach to exploring and understanding emotional and cognitive issues through visual media. This approach involves viewing and analyzing visual media, such as artworks, photography, video, and images<sup>23,24</sup>. Vision therapy can help people understand their emotional and cognitive state. Moreover, it can be used to treat various mental health problems, such as anxiety, depression, and PTSD<sup>25-27</sup>. The effects of visual healing are presented in Fig. 1.

Figure 1 highlights the potential role of vision therapy in providing emotional support and guidance, aiding individuals in coping with issues like anxiety, depression, and PTSD. In the process of treatment, visual therapy offers an intuitive and perceptual communication way through artistic image processing, enabling



**Figure 1.** Schematic diagram of the role of visual healing.

individuals to better understand and express their emotional and cognitive states. Artistic expressions can evoke emotional resonance and facilitate exploration, understanding, and management of internal emotional and cognitive challenges. However, obstacles to effective visual therapy may include resistance or difficulty adapting to art forms, hindering individuals from deriving therapeutic benefits from images. In addition, individuals' understanding and feelings about specific images are subjective, and there may be differences in interpretation, which may also affect treatment effectiveness.

Research supporting visual healing primarily draws on case studies and observations in treating anxiety, depression, and PTSD. Visual therapy encourages individuals to confront and express their emotional issues, fostering better psychological coping mechanisms. While the effectiveness of visual therapy can vary individually, it generally holds promise for providing emotional support and guidance.

### DL and its application to art image processing and analysis

DL is a machine learning method based on artificial neural networks, which has made breakthroughs in many fields, including natural language processing, computer vision, and speech recognition<sup>28,29</sup>. Unlike convolutional networks, self-attention models can efficiently capture and process long-distance dependencies between sequences, leading to robust classification outcomes<sup>30</sup>.

The standard Transformer framework structure comprises two main components: encoder and decoder. The encoder consists of  $M$  layers stacked sequentially, each encompassing two sublayers: a multi-head self-attention mechanism layer and a feedforward neural network layer<sup>31–33</sup>. These layers incorporate residual connection and layer normalization, as shown in Eq. (1).

$$\text{Sublayer}_o = \text{LayerNorm}(x + \text{Sublayer}(x)) \quad (1)$$

$x$  refers to the input of any sublayer; *LayerNorm* means the layer normalization operation, *Sublayer*( $x$ ) represents the function operation performed by the sublayer;  $\text{Sublayer}_o$  denotes the final result of the output of the sublayer.

When normalizing, the mean  $\mu$  and variance  $\sigma$  are calculated by entering the data. The inputs for each dimension are transformed by the same standardized operation, as expressed in Eqs. (2–3).

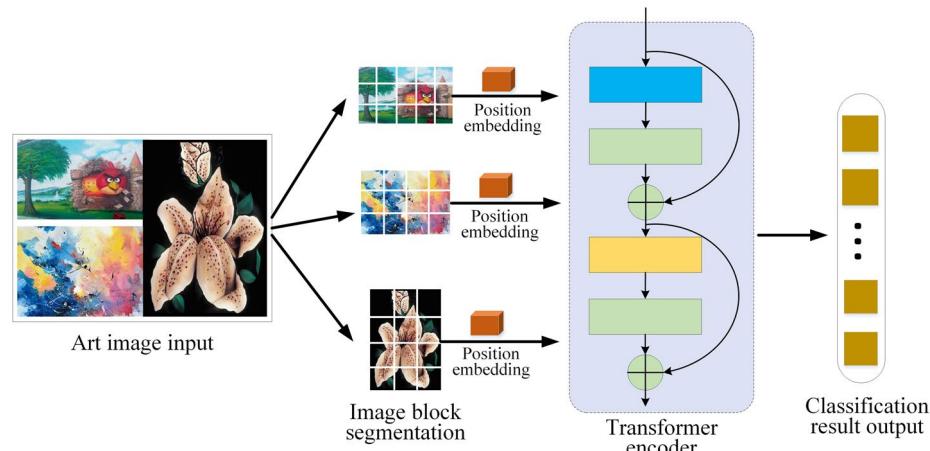
$$\mu = \frac{1}{H} \sum_{i=1}^H x_i \quad (2)$$

$$\sigma = \sqrt{\frac{1}{H} \sum_{i=1}^H (x_i - \mu)^2} \quad (3)$$

$H$  stands for the number of nodes in the hidden layer of a layer;  $x_i$  refers to the input sequence of the layer-represented vector on the  $i$ th image block.

The decoder is similar to an encoder and also requires residual connection and layer normalization of the sublayer<sup>34</sup>.

In the context of artistic image processing, the Vision Transformer (ViT) model uses a Transformer encoder structure to decode collected art image information, generating descriptive features to recognize art image content<sup>35,36</sup>. The framework for applying the ViT model to the analysis of artistic image processing is displayed in Fig. 2.



**Figure 2.** Schematic diagram of the framework of ViT model applied to art image processing analysis.

Figure 2 reveals that in the ViT model, the Transformer encoder performs a trainable linear projection by introducing a convolution operation. Here, the three sub-images in Fig. 2 are simultaneously input into the network as a single input, effectively capturing the global information and local details in the image, thus enhancing image processing outcomes. The sequence of sub-images can significantly influence the results of artistic image processing. Altering the sequence may highlight different local features, thereby impacting the network's interpretation and processing of images. For instance, certain sequences might emphasize texture information more prominently, while others prioritize color distribution. Moreover, different sub-image sequences may affect the network's perception of the overall image structure. Some sequences may be more conducive to the overall organization and layout of images captured by the network, while others might lead it to focus excessively on local details at the expense of the global structure. In addition, rearranging the sequence sub-images can yield varying classification outcomes. The specific ordering may either facilitate or hinder the network's ability to identify specific artistic styles or objects, thus affecting the final image classification results.

Each image block is mapped into a vector with dimension size  $P^2 \times C$  while keeping the number of image blocks  $N$  constant. Then,  $N$  can be expressed as:

$$N = \left\lfloor \frac{h + 2p - P + s}{P - s} + 1 \right\rfloor \times \left\lfloor \frac{w + 2p - P + s}{P - s} + 1 \right\rfloor P^2 \times C \quad (4)$$

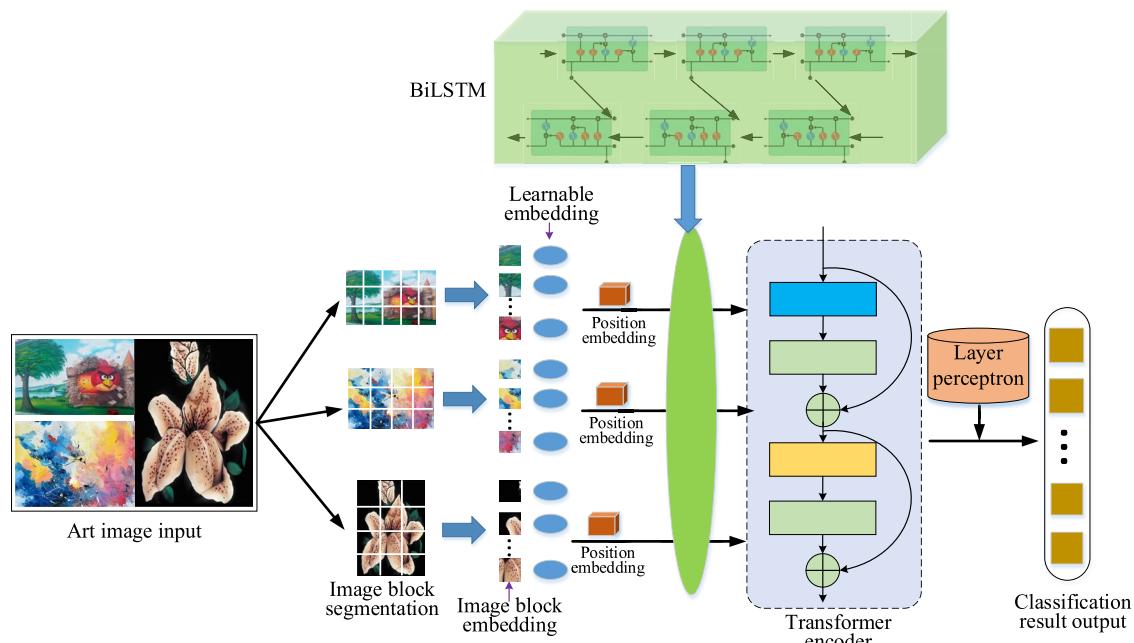
$C$  means the number of channels;  $P \times P$  indicates the size of the image block;  $s$  refers to the overlapping width of adjacent picture blocks;  $p$  stands for the fill size of the input preprocessed picture;  $w$  and  $h$  denote the convolution operation's sliding step size.  $N$  and  $P$  represent the number of blocks.

However, ViT models often suffer from feature loss at the boundaries between adjacent image blocks. To address this challenge and enhance the recognition and classification accuracy of artistic images, an improved ViT model called OSViT is proposed. OSViT integrates overlapping recognition and classification strategies, specifically designed for the processing and classification of artistic images.

### Construction of art image processing and classification recognition model based on DL

Under the vision of SC, to preserve the sequence information in artistic images, this study improves the ViT model and designs the OSViT algorithm. OSViT comprises five main modules: image block embedding, learnable embedding, position embedding, Transformer encoder, and layer perceptron. At the same time, the BiLSTM algorithm is integrated to enable the model to capture the sequence information from the image during processing and classification, while OSViT is responsible for processing the information of the image block. The BiLSTM algorithm plays a crucial role in this process. The order of image blocks in artistic images often carries important semantic information, such as the arrangement order of objects and contextual information. This capability allows the model to accurately understand the semantic structure of images. By leveraging the BiLSTM algorithm, the enhanced model effectively utilizes temporal sequence information within images, thereby improving the accuracy and robustness of image recognition and classification. The model's overall framework is depicted in Fig. 3.

Figure 3 shows that in this model algorithm, the image block sequence is divided into a fixed number of patches to facilitate the model's standardized processing and training. A learnable classification vector  $z_0^0 = x_{class}$



**Figure 3.** Schematic diagram of art image processing and classification recognition model based on OSViT-BiLSTM.

is added to the sequence of image blocks. The final feature of this vector output by the Transformer encoder is used as a representation of the image category. Vector  $z_0^0$  can be regarded as the category information that all other image patches are looking for. The added learnable category embedding equation is:

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^N E] \quad (5)$$

$x_{class}$  represents the vector of the classification category;  $x_p^1, x_p^2, \dots, x_p^N$  refers to the image block in the input image after preprocessing;  $E$  means the weight of the linear projection.

The vector dimension of positional encoding needs to be consistent with the dimension of the image block embedding, and the positional encoding vector can be calculated using cosine and sine functions of different frequencies. Thus, the equation for adding positional encoding information to the serialized art image block is as follows.

$$E_{pos}(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (6)$$

$$E_{pos}(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (7)$$

$pos$  stands for the position of the current image block in the input image;  $i$  refers to the index of each value in the vector;  $d_{model}$  represents the dimension of the art image block.

Then, the calculated positional encoding vector and the image block embedding vector sequence are added as input to the encoder, as follows.

$$Z_0 = Z + E_{pos} \quad (8)$$

$Z$  refers to the value after the linear projection of the original image;  $E_{pos}$  denotes the position information vector. This vector determines the position of the current image block in the sequence of vectors composed of image blocks.  $Z_0$  represents the sequence of image block vectors that add position information as input to subsequent Transformer encoders. After embedding the position information of the original image, the BiLSTM module is introduced to process the sequence information related to the artistic image. This can include the time series changes of objects or colors in the image, etc. BiLSTM is designed to capture sequential patterns in images.

In the Transformer encoder module, the features of the input encoder are first layer-normalized, and the feature  $M_0$  is adjusted.

$$M_0 = \frac{Z_0 - \mu}{\sqrt{\sigma + \varepsilon}} \cdot \gamma + \beta \quad (9)$$

$Z_0$  means the input vector of the encoder;  $\mu$  and  $\sigma$  refer to its mean and variance, respectively.  $\varepsilon$  is a very small number set to avoid a denominator of zero, and its value is set to  $10^{-8}$ .  $\gamma$  and  $\beta$  represent the parameter vectors of scaling and translation, respectively, consistent with the representation vector dimensions of the image block of the input encoder. Then, the features after layer normalization adjustment features are calculated by multi-head self-attention.

$$head_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (10)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1 \dots head_h) W^o \quad (11)$$

$Q$ ,  $K$ , and  $V$  represent the query, key, and value of the attention calculation process, respectively. During the linear calculation of the multi-head self-attention, parameters between the heads  $head_i$  cannot be shared. The linear transformation of the  $Q$ ,  $K$ , and  $V$  eigenvectors is calculated using matrices  $W_i^Q$ ,  $W_i^K$ ,  $W_i^V$ . The calculation result of the contraction point product of  $h$  times is spliced. Finally, the final result of the multi-head self-attention is obtained through a linear transformation.  $W^o$  refers to the weight matrix of the linear transformation of the multi-head attention calculation.

The layer perceptron module sends the input features into the linear layer for linear transformation. Subsequently, activation is performed using the Gaussian Error Linear Units (GELUs) activation function<sup>37</sup>.

$$GELU = x\Phi(x) \quad (12)$$

$x$  refers to the linear layer's output feature;  $\Phi(x)$  means the probability function of the normal distribution.

The output result of the Dropout layer is connected to the calculation result of the first residual connection in the encoder, which represents the layer perceptron module's final output.

The proposed OSViT-BiLSTM model algorithm's pseudo-code is exhibited in Algorithm 1.

---

```

1 Start
2 Input: art image data
3 Output: artistic image processing and classification recognition results
4 # Define OSViT model
5 class OSViTModel(nn.Module):
6     def __init__(self, ...): # Define parameters for OSViT
7         def forward(self, image):
8             # Define BiLSTM model
9             class BiLSTMModel(nn.Module):
10                 def __init__(self, ...): # Define parameters for BiLSTM
11                     def forward(self, sequence):
12                         # Define integrated model
13                         class ArtImageModel(nn.Module):
14                             def __init__(self, osvit_params, lstm_params, transformer_params, num_classes):
15                                 super(ArtImageModel, self).__init__()
16                                 self.osvit = OSViTModel(osvit_params)
17                                 self.bilstm = BiLSTMModel(lstm_params)
18                             def forward(self, image, sequence):
19                                 osvit_output = self.osvit(image)
20                                 bilstm_output = self.bilstm(sequence)
21 End

```

---

**Analysis of the classification and application of art image processing in visual healing in SC**  
 The above DL technology is utilized to process and identify art images in SC, enabling their classification and application across various fields to achieve visual healing effects. The classification of visual healing applications after art image classification recognition is given in Fig. 4.

In Fig. 4, after processing the art images, visual healing application scenarios can be categorized into several aspects, including art image display, SC public facilities, and the application of visual healing in transportation. The art image display module aims to enhance citizens' aesthetic literacy and emotional experience by showcasing outstanding artworks in urban public places, such as park display boards and public newsstands. The application of visual healing in SC public facilities focuses on improving citizens' physical and mental health through visual healing in public facilities, such as public street electrical boxes and community garbage cans. The visual healing



**Figure 4.** Schematic diagram of visual healing application classification after art image classification recognition.

application module in SC transportation incorporates artistic images in traffic scenes to reduce traffic-related stress and anxiety among citizens.

## Experimental design and performance evaluation

### Datasets collection and experimental environment

The data source is the WikiArt dataset (<https://www.wikiart.org/>). The collection contains paintings, sculptures, photography, and other art forms, totaling approximately 200,000 pieces<sup>38</sup>. A total of 187,000 high-quality art images covering various themes in the WikiArt dataset are selected as experimental subjects and randomly divided into training and test sets in an 8:2 ratio. First, the Linux operating system servers or cloud virtual machines are installed, and TensorFlow is used to install DL frameworks and necessary Python packages and tool libraries, such as NumPy<sup>39</sup>. Second, GPU is employed to accelerate the model training speed. Finally, Git and Docker are utilized to manage experimental data and results.

### Parameters setting

When using the OSViT algorithm to process artistic images, all input images are pre-processed into three channels and the size is  $224 \times 224$  pixels. The convolution kernel size of the image block embedding layer is 16, with a stride of 16, dividing the image into 196 image blocks. The image block vector dimension for the input encoder module is 768, and the encoder's Q, K, and V sizes are all 3. The image block vector dimension is 2304, and the number of heads for the multi-head self-attention is 8. The output of the first linear layer in the multilayer perceptron module is 3072, while the second linear layer's output, activated by GELU, is 7680. The number of self-attention submodules is 12. The initial learning rate is set to 0.01, with a batch input of 8 images per training session, and the training runs for a total of 70 cycles.

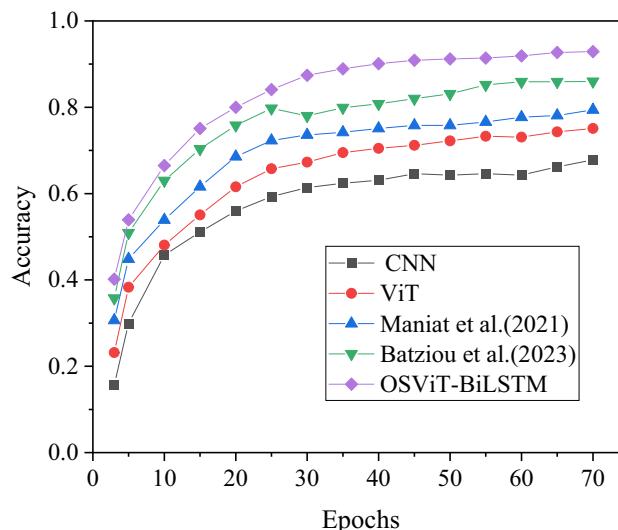
### Performance evaluation

The proposed OSViT-based art image processing and classification recognition model algorithm is analyzed in comparison with ViT, Convolutional Neural Network (CNN) (Naseer et al.<sup>40</sup>, Maniat et al.<sup>19</sup>, and Batziou et al.<sup>21</sup> based on accuracy, precision, recall, and F1 score. Subsequently, the accuracy of image classification recognition is evaluated. The results are revealed in Figs. 5, 6, 7 and 8.

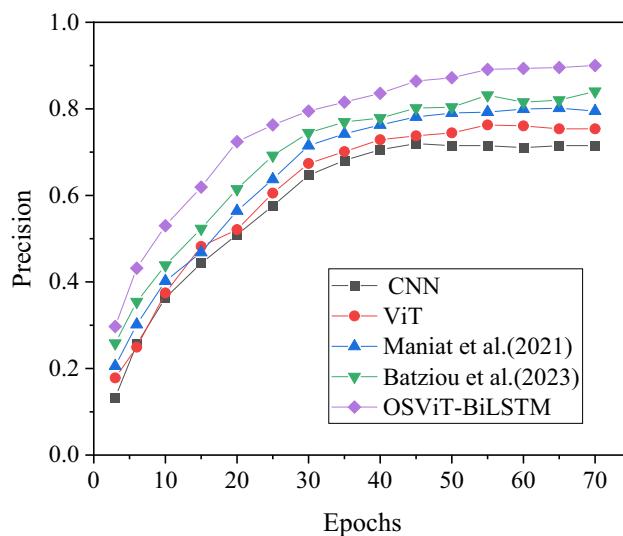
Figures 5, 6, 7 and 8 denote that in this comparative analysis, each algorithm shows an initial increase in recognition accuracy followed by stabilization with increased training cycles. Specifically, the proposed model algorithm achieves an accuracy of 92.90% in classifying and recognizing artistic images, which is at least 6.90% higher than other models proposed by scholars. Regarding F1 score, precision, and recall, the proposed algorithm reaches 76.30, 90.02, and 84.50%, respectively. When ranked by accuracy, precision, recall, and F1 score from highest to lowest, the results indicate the superiority of the OSViT-BiLSTM algorithm > Batziou et al.<sup>21</sup> > Maniat et al.<sup>19</sup> > ViT > CNN. This demonstrates that under the vision of SC, the artistic image processing and classification recognition model based on OSViT-BiLSTM developed here can effectively predict and classify artistic images, achieving accurate recognition and classification outcomes.

In addition, the recognition accuracy of this model algorithm is analyzed through ablation experiments, and the proposed OSViT-BiLSTM algorithm is compared with OSViT, ViT-BiLSTM, and ViT, and its accuracy is suggested in Fig. 9.

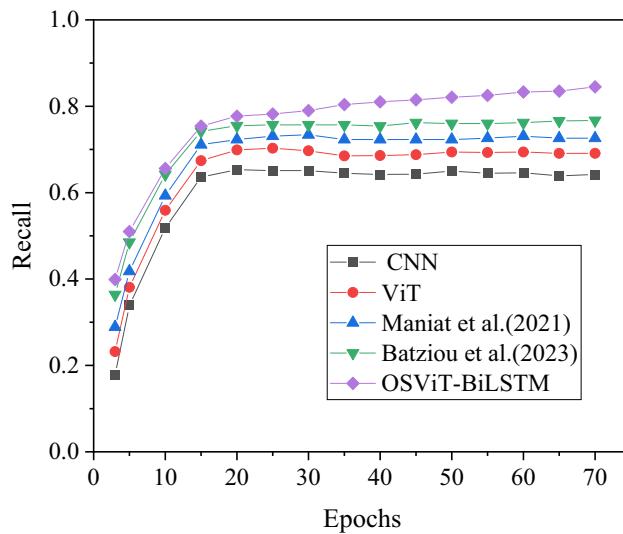
Figure 9 illustrates the results of the ablation experiment, showing that as the number of training iterations increases, the recognition accuracy of each algorithm initially rises and then stabilizes. Among them, the accuracy of the classification and recognition of artistic images by the proposed algorithm reaches 92.9%, which is



**Figure 5.** Accuracy result of art image processing and classification recognition under diverse algorithms.



**Figure 6.** Precision result of art image processing and classification recognition under different algorithms.



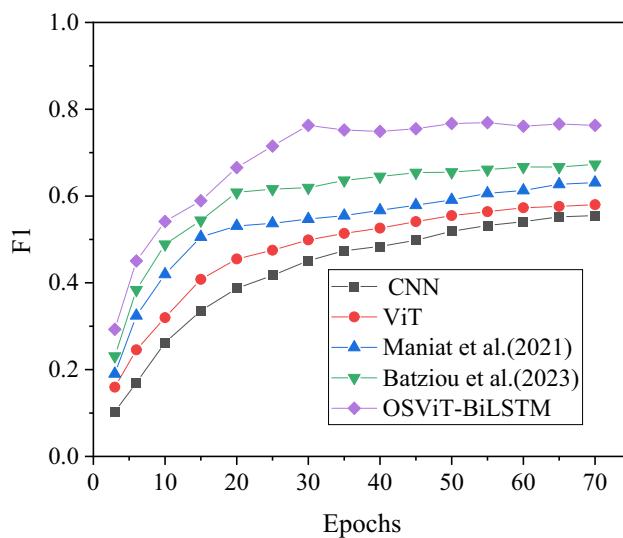
**Figure 7.** Recall of art image processing and classification recognition with various algorithms.

markedly improved compared with OSViT, ViT-BiLSTM, and ViT algorithms. Among them, the ViT algorithm's identification accuracy is only 71.09%. Therefore, the proposed OSViT-BiLSTM algorithm can accurately process and classify artistic images in the ablation experiment.

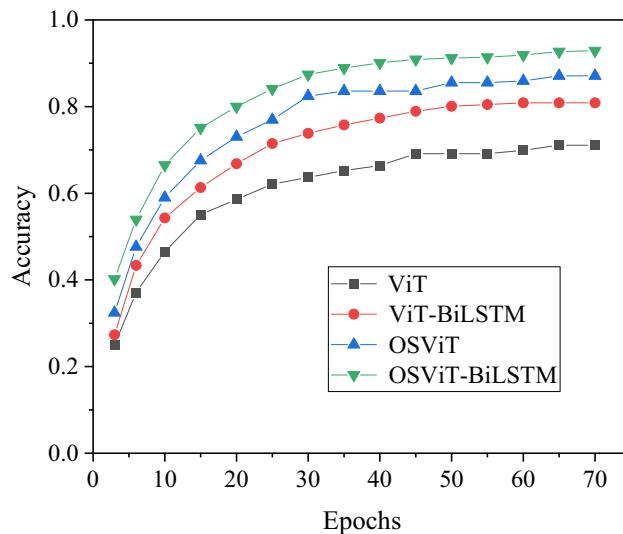
Furthermore, the effects of visual healing and pressure release of processed and classified artistic images in different scenes and cultural levels are analyzed. 150 test users are randomly selected to rate the visual healing effects of art images in different scenes using the Likert 5-point scale. Among them, very satisfied = 5 points, satisfied = 4 points, general = 3 points, dissatisfied = 2 points, very dissatisfied = 1 point. This questionnaire test has obtained the informed consent of the participants. The results are plotted in Figs. 10 and 11.

In Fig. 10, when the processed artistic images are applied to various scenes, such as public scenes, vehicles, homes, public infrastructure, and workplaces, more than 90% of users score 4 or above. It means that the healing effect of artistic images in different scenes is satisfactory, which brings pressure release and psychological relief effects. This shows that the proposed algorithm can effectively deal with and accurately identify artistic processing. Thus, it can positively impact users' mental health in diverse scenarios, and provide a potentially positive effect for pressure release, with a visual healing effect.

Figure 11 exhibits the varying attitudes toward artistic image therapy based on respondents' educational levels. Among primary school students, 11.36% express dissatisfaction with artistic image therapy. In contrast, less than 5% of junior high school students are dissatisfied. Over 93% of users with undergraduate education support the therapy, with no expressions of dissatisfaction. Generally, users with undergraduate education levels demonstrate high satisfaction with the therapy, while users with junior high school education levels are also satisfied. The



**Figure 8.** F1 score of art image processing and classification recognition under different algorithms.



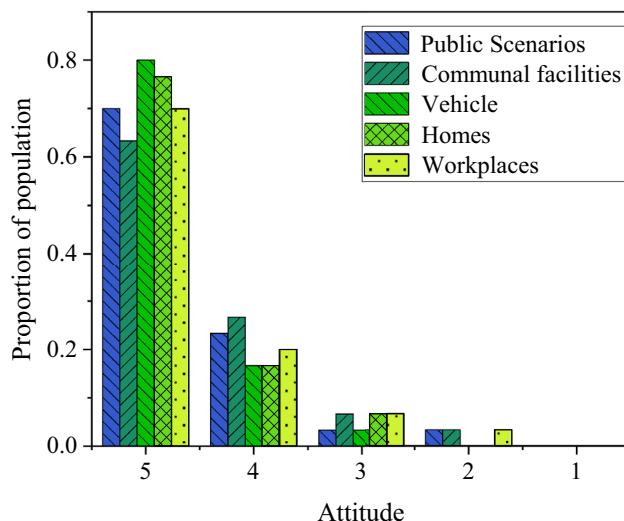
**Figure 9.** Accuracy value result of artistic image processing and classification recognition in ablation experiment.

relatively low satisfaction in primary school may be related to students' understanding and cognitive level of cultural differences at this stage. Users with undergraduate education levels may be more inclined to actively accept and understand the healing effect of artistic images, which can positively impact stress relief and mental health improvement in daily life. Thus, enhancing educational levels could potentially increase awareness of stress management techniques, thereby boosting satisfaction with artistic image therapy.

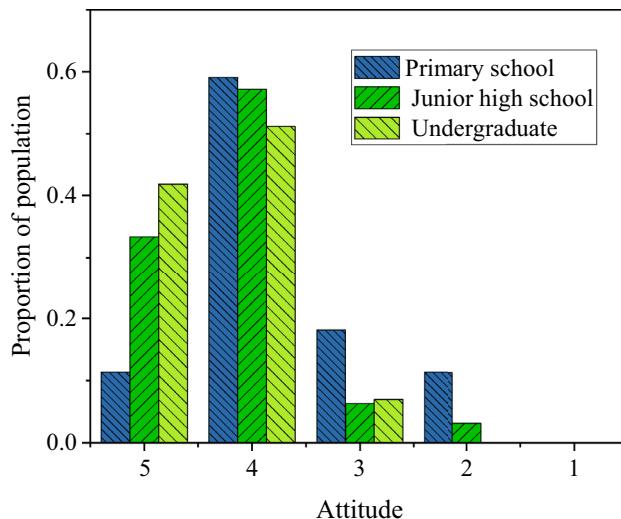
## Discussion

The processing effect of the proposed OSViT model algorithm on artistic images is analyzed. It can be found that compared with the image classification recognition accuracy of other existing models, including ViT, CNN, Maniat et al.<sup>19</sup>, the proposed model algorithm achieves 89.80% accuracy in art image classification recognition, remarkably exceeding the model algorithm proposed by other scholars. It suggests that the OSViT-BiLSTM algorithm-based art image processing and classification model can accurately predict and classify art images, which can contribute to the accurate recognition and classification of art images.

Furthermore, when applying processed art images to various public settings such as public spaces, infrastructure, and transportation, more than 90% of users rate the visual healing effects of these images as satisfactory. This demonstrates that the proposed model algorithm effectively processes and accurately identifies artistic images, thereby achieving user satisfaction and promoting visual healing effects across different



**Figure 10.** Visual healing effect after artistic processing and application to different scenes.



**Figure 11.** After artistic image processing, it is applied to the visual healing effect map of users with different educational levels.

scenarios<sup>41</sup>. Hence, within the scope of SC, this study establishes a feasible and practical application scheme for visual healing in diverse settings following art image processing and recognition.

## Conclusion

To address the pressures and challenges posed by urbanization on individuals' lives and work, an art image processing and classification recognition model based on the OSViT-BiLSTM algorithm is developed. The processed art images are applied to different scenes to analyze their visual healing effects. The findings underscore that the model algorithm proficiently processes and accurately identifies art images, facilitating precise recognition and classification under the SC framework. Moreover, art images processed through the OSViT-BiLSTM model algorithm exhibit substantial visual healing benefits in public spaces, infrastructure, and transportation settings. Thus, this study has the value of promoting and applying diverse art images in public settings. Nonetheless, there are limitations, such as the relatively small sample size of test users. Future research could enhance understanding by investigating the impact of different image processing technologies and model parameters on the recognition and classification of artistic images across various contexts through expanded sample sizes. Additionally, exploring how cultural differences influence perceptions of the healing effects of artistic images represents a valuable direction for further investigation.

## Ethics approval

This article does not contain any studies with human participants or animals performed by any of the authors. All methods were performed in accordance with relevant guidelines and regulations.

## Data availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The data will be requested to be available by Guangfu Qu via email qgf1974@163.com.

Received: 27 October 2023; Accepted: 19 July 2024

Published online: 22 July 2024

## References

1. Ghazal, T. M. *et al.* IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet* **13**(8), 218 (2021).
2. Zhu, H., Shen, L. & Ren, Y. How can smart city shape a happier life? The mechanism for developing a happiness driven smart city. *Sustain. Cities Soc.* **80**, 103791 (2022).
3. Radu, L. D. Disruptive technologies in smart cities: A survey on current trends and challenges. *Smart Cities* **3**(3), 1022–1038 (2020).
4. Sik, D. From lay depression narratives to secular ritual healing: An online ethnography of mental health forums. *Cult. Med. Psychiatry* **45**(4), 751–774 (2021).
5. Kohrt, B. A., Ottman, K., Panter-Brick, C., Konner, M. & Patel, V. Why we heal: The evolution of psychological healing and implications for global mental health. *Clin. Psychol. Rev.* **82**, 101920 (2020).
6. Adjapong, E. & Levy, I. Hip-hop can heal: Addressing mental health through hip-hop in the urban classroom. *New Educ.* **17**(3), 242–263 (2021).
7. Basch, C. H., Donelle, L., Fera, J. & Jaime, C. Deconstructing TikTok videos on mental health: Cross-sectional, descriptive content analysis. *JMIR Format Res.* **6**(5), e38340 (2022).
8. Bensaoud, A. & Kalita, J. Deep multi-task learning for malware image classification. *J. Inf. Secur. Appl.* **64**, 103057 (2022).
9. Ma, W., Tu, X., Luo, B. & Wang, G. Semantic clustering based deduction learning for image recognition and classification. *Pattern Recognit.* **124**, 108440 (2022).
10. Wang, Y., Huang, R., Song, S., Huang, Z. & Huang, G. Not all images are worth  $16 \times 16$  words: Dynamic transformers for efficient image recognition. *Adv. Neural Inf. Process. Syst.* **34**, 11960–11973 (2021).
11. Freeman, R. C. Jr *et al.* Promoting spiritual healing by stress reduction through meditation for employees at a veterans hospital: A CDC framework-based program evaluation. *Workplace Health Safety* **68**(4), 161–170 (2020).
12. Hass-Cohen, N., Bokoch, R., Goodman, K. & Conover, K. J. Art therapy drawing protocols for chronic pain: Quantitative results from a mixed method pilot study. *Arts Psychother.* **73**, 101749 (2021).
13. Bowen-Salter, H. *et al.* Towards a description of the elements of art therapy practice for trauma: A systematic review. *Int. J. Art Ther.* **27**(1), 3–16 (2022).
14. Lee, H., Kim, E. & Yoon, J. Y. Effects of a multimodal approach to food art therapy on people with mild cognitive impairment and mild dementia. *Psychogeriatrics* **22**(3), 360–372 (2022).
15. Julianino, C., Nathania, M. P., Hendarti, R., Darmadi, H. & Suryawinata, B. A. The development of virtual healing environment in VR platform. *Proc. Comput. Sci.* **216**, 310–318 (2023).
16. Skorburg, J. A., O'Doherty, K. & Friesen, P. Persons or data points? Ethics, artificial intelligence, and the participatory turn in mental health research. *Am. Psychol.* **79**(1), 137 (2024).
17. Mitro, N. *et al.* AI-enabled smart wristband providing real-time vital signs and stress monitoring. *Sensors* **23**(5), 2821 (2023).
18. Ayana, G., Park, J., Jeong, J. W. & Choe, S. W. A novel multistage transfer learning for ultrasound breast cancer image classification. *Diagnostics* **12**(1), 135 (2022).
19. Maniat, M., Camp, C. V. & Kashani, A. R. Deep learning-based visual crack detection using google street view images. *Neural Comput. Appl.* **33**(21), 14565–14582 (2021).
20. Wang, X., Chen, Z., Ren, J., Chen, S. & Xing, F. Object status identification of X-ray CT images of microcapsule-based self-healing mortar. *Cement Concr. Compos.* **125**, 104294 (2022).
21. Batziou, E., Ioannidis, K., Patras, I., Vrochidis, S. & Kompatiari, I. Artistic neural style transfer using CycleGAN and FABEMD by adaptive information selection. *Pattern Recognit. Letter* **165**, 55–62 (2023).
22. Tehsin, S., Kausar, S., Jameel, A., Humayun, M. & Almofarreh, D. K. Satellite image categorization using scalable deep learning. *Appl. Sci.* **13**(8), 5108 (2023).
23. McCrory, A., Best, P. & Maddock, A. The relationship between highly visual social media and young people's mental health: A scoping review. *Child. Youth Serv. Rev.* **115**, 105053 (2020).
24. Yazdavar, A. H. *et al.* Multimodal mental health analysis in social media. *Plos one* **15**(4), e0226248 (2020).
25. Dobersek, U. *et al.* Meat and mental health: A systematic review of meat abstention and depression, anxiety, and related phenomena. *Crit. Rev. Food Sci. Nutrit.* **61**(4), 622–635 (2021).
26. Meeks, K., Peak, A. S. & Dreihaus, A. Depression, anxiety, and stress among students, faculty, and staff. *J. Am. College Health* **71**(2), 348–354 (2023).
27. Malaeab, D. *et al.* Problematic social media use and mental health (depression, anxiety, and insomnia) among Lebanese adults: Any mediating effect of stress? *Perspect. Psychiatr. Care* **57**(2), 539–549 (2021).
28. Su, C., Xu, Z., Pathak, J. & Wang, F. Deep learning in mental health outcome research: A scoping review. *Transl. Psychiatry* **10**(1), 116 (2020).
29. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**(7914), 501–506 (2022).
30. AprilPyone, M. & Kiya, H. Privacy-preserving image classification using an isotropic network. *IEEE MultiMed.* **29**(2), 23–33 (2022).
31. Liu, X., Wang, L. & Han, X. Transformer with peak suppression and knowledge guidance for fine-grained image recognition. *Neurocomputing* **492**, 137–149 (2022).
32. Zhang, Q. & Yang, Y. B. Rest: An efficient transformer for visual recognition. *Adv. Neural Inf. Process. Syst.* **34**, 15475–15485 (2021).
33. Kamal, M. B. *et al.* An innovative approach utilizing binary-view transformer for speech recognition task. *Comput. Mater. Contin.* **72**(3), 5547–5562 (2022).
34. Rao, Y. *et al.* Dynamiccvit: Efficient vision transformers with dynamic token sparsification. *Adv. Neural Inf. Process. Syst.* **34**, 13937–13949 (2021).
35. Rajan, K., Zielesny, A. & Steinbeck, C. DECIMER 1.0: Deep learning for chemical image recognition using transformers. *J. Cheminf* **13**(1), 1–16 (2021).
36. Yuan, F., Zhang, Z. & Fang, Z. An effective CNN and transformer complementary network for medical image segmentation. *Pattern Recognit.* **136**, 109228 (2023).

37. Sun, L., Zhao, G., Zheng, Y. & Wu, Z. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–14 (2022).
38. Tashu, T. M., Hajiyeva, S. & Horvath, T. Multimodal emotion recognition from art using sequential co-attention. *J. Imaging* **7**(8), 157 (2021).
39. Goel, P. Realtime object detection using tensorflow an application of ML. *Int. J. Sustain. Dev. Comput. Sci.* **3**(3), 11–20 (2021).
40. Naseer, I. *et al.* Performance analysis of state-of-the-art CNN architectures for luna16. *Sensors* **22**(12), 4426 (2022).
41. Vaartio-Rajalin, H., Santamäki-Fischer, R., Jokisalo, P. & Fagerström, L. Art making and expressive art therapy in adult health and nursing care: A scoping review. *Int. J. Nurs. Sci.* **8**(1), 102–119 (2021).

## Acknowledgements

This work was supported by 2024 Shandong Province Art Science Key Project of “Application Research of AIGC Enabling Archaeological Documentary Creation”.

## Author contributions

Guangfu Qu: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation Qian Song: Conceptualization, methodology, software, validation, formal analysis, writing—review and editing Ting Fang: visualization, supervision, project administration, funding acquisition.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.Q. or T.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024