



OPEN ACCESS

EDITED BY
Deepika Koundal,
University of Petroleum and Energy
Studies, India

REVIEWED BY
Arvind Dhaka,
Manipal University Jaipur, India
Mohit Mittal,
Institut National de Recherche en
Informatique et en Automatique
(INRIA), France
Tariq Ahmad,
Guilin University of Electronic
Technology, China

*CORRESPONDENCE
Kai Cheng
✉ chengkai7300@163.com

RECEIVED 06 December 2023
ACCEPTED 28 March 2024
PUBLISHED 17 April 2024

CITATION
Cheng K (2024) Prediction of emotion
distribution of images based on weighted
 K -nearest neighbor-attention mechanism.
Front. Comput. Neurosci. 18:1350916.
doi: 10.3389/fncom.2024.1350916

COPYRIGHT
© 2024 Cheng. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Prediction of emotion distribution of images based on weighted K -nearest neighbor-attention mechanism

Kai Cheng*

School of Artificial Intelligence, Xidian University, Xi'an, China

Existing methods for classifying image emotions often overlook the subjective impact emotions evoke in observers, focusing primarily on emotion categories. However, this approach falls short in meeting practical needs as it neglects the nuanced emotional responses captured within an image. This study proposes a novel approach employing the weighted closest neighbor algorithm to predict the discrete distribution of emotion in abstract paintings. Initially, emotional features are extracted from the images and assigned varying K -values. Subsequently, an encoder-decoder architecture is utilized to derive sentiment features from abstract paintings, augmented by a pre-trained model to enhance classification model generalization and convergence speed. By incorporating a blank attention mechanism into the decoder and integrating it with the encoder's output sequence, the semantics of abstract painting images are learned, facilitating precise and sensible emotional understanding. Experimental results demonstrate that the classification algorithm, utilizing the attention mechanism, achieves a higher accuracy of 80.7% compared to current methods. This innovative approach successfully addresses the intricate challenge of discerning emotions in abstract paintings, underscoring the significance of considering subjective emotional responses in image classification. The integration of advanced techniques such as weighted closest neighbor algorithm and attention mechanisms holds promise for enhancing the comprehension and classification of emotional content in visual art.

KEYWORDS

image emotions, classification, weighted closest neighbor algorithm, emotional features, abstract paintings

1 Introduction

Image data are essentially used for transferring information. The amount of picture data is even increasing at an exponential speed owing to the advent of the Internet (Cetinic and She, 2022; Zou et al., 2023). Because of the fast-paced nature of modern society, people's ability to extract information from photos is also accelerating, necessitating more accuracy and efficiency in identifying image data on the network. Based on this necessity, an effective image processing technique that makes use of computer vision is required for humans to manage and use picture data more effectively.

Sentiment analysis, often called opinion mining, is the process of using natural language processing, text analysis, computational linguistics, and biometrics to systematically unpack subjective information and emotional states. The notion was initially introduced by Yang et al. (2023). Sentiment analysis has gained significant economic and societal significance in the last several years and has been applied extensively in the domains of opinion monitoring (Chen et al., 2023), topic inference (Ngai et al., 2022), and comment analysis and decision-making (Bharadiya, 2023). For monitoring public opinion, the government can make timely policy interventions and accurately determine the direction of public opinion. When it comes to product recommendations, merchants can better understand user needs and suggestions by gauging user satisfaction with product evaluations and enhancing product quality. In the finance domain, trending financial topics can even be used to predict stock direction. Furthermore, sentiment analysis is frequently used for various tasks involving natural language processing. To increase the accuracy of the system, more exact terms for sentiment expression are chosen for machine translation (Chan et al., 2023) by evaluating the sentiment tendency of the input text. The pixel density extraction of the image information is shown in Figure 1.

Various classification techniques will be broken down into different levels for the sentiment analysis task: output results will categorize the methods into sentiment intensity classification and sentiment polarity classification; granularity of the processed text will divide them into three research levels: word level, sentence level, and chapter level; research methodology will separate them into unsupervised learning, semi-supervised learning, and supervised learning, and so on. The majority of the conventional sentiment classification algorithms employ manually created feature selection techniques for feature extraction, such as the maximum entropy model (Chandrasekaran et al., 2022), plain Bayes (Wang et al., 2022), support vector machines (Zhao et al., 2021a), and so on. However, these techniques have limitations, such as being labor-intensive, time-consuming, and hard to train. As a result, they are not well-suited for use in the current large-scale application scenarios.

With advancements in machine learning, research efforts (Milani and Fraternali, 2021) led to the development of deep learning methods that give neural networks a hierarchical structure. This development subsequently resulted in an explosion of deep learning research. Feature learning, at the heart of deep learning, uses hierarchical networks to convert unprocessed input into more abstract and higher-level feature information. With its superior learning capacity to optimize automated feature extraction, deep learning has produced remarkable research achievements in recent years in the domains of speech recognition, picture processing, and natural language processing. The application of deep learning techniques to text sentiment analysis has gained popularity as a natural language processing study area. Among these techniques, Song et al. (2021) used a convolutional neural network to classify text emotion for the first time, and the results were superior to those of conventional machine learning techniques.

The study of human eyesight is where attention mechanism first emerged. According to cognitive science, humans have a tendency to ignore other observable information in favor of focusing on a certain portion of the information based on the

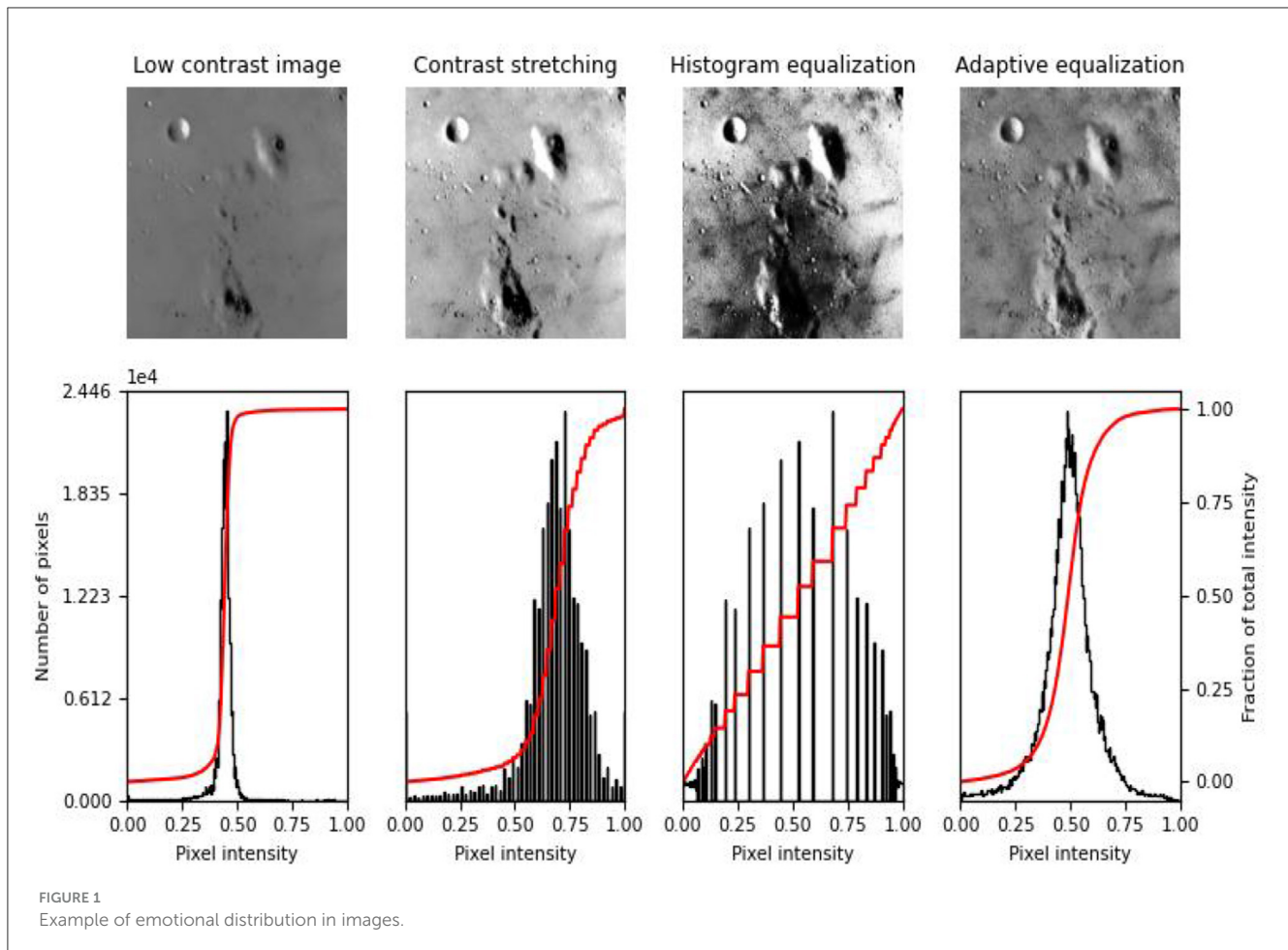
demand imposed by the information processing bottleneck. The primary objective of attention mechanism is to efficiently separate valuable information from a vast quantity of data. To understand the word dependencies inside the phrase and grasp the internal structure of the sentence, the self-attention mechanism—a unique form of attention mechanism—is incorporated into the sentiment classification job. To establish an accurate and efficient technique for sentiment analysis based on deep learning technology and self-attention mechanism, this study examines the present technical issues in the field of sentiment analysis from the standpoint of the real demands of sentiment analysis.

2 Related studies

Natural language processing has attracted extensive research attention (McCormack and Lomas, 2021) because it introduced the idea of sentiment analysis. There are three prominent methods for conducting sentiment analysis at present: the sentiment dictionary approach, the classical machine learning approach, and the deep learning approach.

Experts must annotate the sentiment polarity of the text's terms in order for researchers to perform sentiment analysis based on sentiment dictionary. Based on semantic rules and sentiment dictionary, researchers compute the text's sentiment score and determine the sentiment tendency. Among these researchers, Toisoul et al. (2021) demonstrated positive findings on a multi-domain dataset by expanding the domain-specific vocabulary by extracting subject terms from the corpus using latent Dirichlet allocation (LDA) modeling based on the pre-existing sentiment lexicon. Peng et al. (2022) used the point mutual information (PMI) technique to assess the similarity of adjectives in WordNet. The polar semantics (ISA) approach was then used to generate numerous fixed sentence constructions in order to examine the target text sentiment tendency. To create a Chinese microblogging sentiment dictionary, Liu et al. (2021a) first identified microblogging sentences using information entropy and then filtered network sentiment terms using the sentiment-oriented pointwise mutual information (SO-PMI) method.

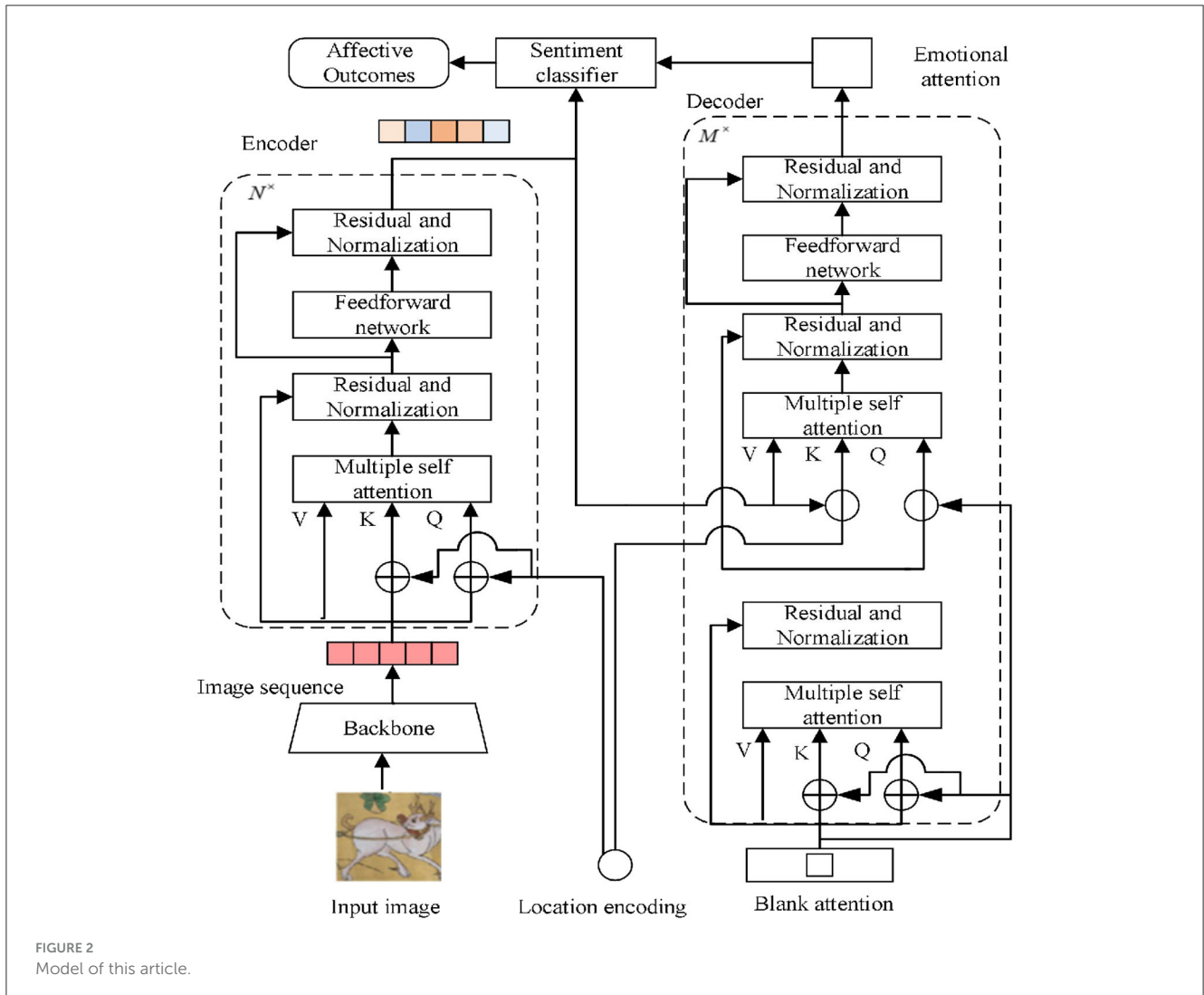
Ding et al. (2021) introduced the idea of the primary word and used weight priority calculations to determine the text's semantic inclination degree. These developments paved the way for accomplishing more difficult sentiment analysis tasks. The approach based on sentiment dictionary has the benefit of being more accurate in classifying text at the word or phrase level. However, the system migration is not good, and the sentiment dictionaries are often geared to certain domains. These days, one of the most popular techniques for sentiment analysis is classical machine learning-based techniques. Using simple bag-of-words features from a movie review dataset, Yang et al. (2021) was the first to use machine learning techniques to the sentiment binary classification issue and produced superior experimental outcomes. Utilizing Twitter comments as test data, Roy et al. (2023) classified emotions into six categories—happiness, sadness, disgust, fear, surprise, and anger—and employed plain Bayes for text sentiment analysis. The data were processed with consideration for lexical and expression features, leading to a high classification accuracy. To



address the sentiment classification problem, [Sahoo et al. \(2021\)](#) merged a genetic algorithm with simple Bayes, and the results of the experiments indicated that the combined model outperformed the individual models. To extract rich sentiment data and include them in the basic feature model, [Liu et al. \(2021b\)](#) used machine learning techniques with numerous rules, which increased the classification result in microblog sentiment classification trials. In order to complete the study of sentiment analysis, [Sampath et al. \(2021\)](#) included semantic rules into the support vector machine model. The experiment confirmed that the support vector machine model with the inclusion of semantic rules performed better in the sentiment classification task. Deeper text semantic information is hard to learn, even while machine learning-based techniques enhance the sentiment classification performance and lower the reliance on sentiment lexicon.

Text sentiment analysis based on deep learning has garnered a much interest from academics at both national and international levels due to its superior performance in the fields of picture processing and natural language processing. [Zhang et al. \(2023\)](#) used deep neural network training to create the Collobert and Weston (C&W) model, which was then used to perform well on natural language processing tasks including sentiment classification and lexical annotation. To demonstrate the efficacy of single-layer convolutional neural networks (CNNs) in sentiment classification tasks, [Zhao et al. \(2021b\)](#) combined different sizes of convolutional

kernels with maximum pooling and performed comparison tests on seven datasets. The study employed convolutional neural networks for sentiment analysis tasks. A number of recurrent neural networks, including recurrent neural network (RNN), multiplicative RNN (MRNN), recursive neural tensor network (RNTN), and others, were progressively suggested by [Szubielska et al. \(2021\)](#). The RNTN model, for example, uses a syntactic analysis tree to determine word sentiment and then outputs the sentence's sentiment classification result in the form of word sentiment summation. To tackle the sentiment analysis problem utilizing a long short-term memory (LSTM) network with an expanded gate structure, which increases the model's flexibility, [Li et al. \(2022\)](#) employed Twitter comments as the experimental data. RNNs were utilized by [Zhou et al. \(2023\)](#) to model texts by taking into account their temporal information. [Li et al. \(2023\)](#) achieved outstanding results in a sentiment classification test by modeling utterances using a tree LSTM model to approximate the sentence structure. By segmenting a text according to sentences, obtaining vectors through convolutional pooling operation, and then inputting them into LSTM according to temporal relations to construct a CNN-LSTM model and apply it to the task of sentiment analysis, [Alirezazadeh et al. \(2023\)](#) primarily addressed the issue of temporal and long-range dependencies in a chapter-level text. [Teodoro et al. \(2023\)](#) constructed an experimental minimal convolutional neural



network (EMCNN) model using microblog comments as the experimental data, combining lexical and emoji characteristics. The model produced experimental findings that outperformed the benchmark model's performance.

3 Attention given

We propose an emotion classification method based on the attention mechanism that sets blank attention in the decoder and fuses the output sequence of the encoder to learn the image semantics to guide the model to learn the image emotion more accurately and reasonably via the learning mechanism of the decoder. This method is intended to address the characteristics of small numbers of abstract painting samples and rich image semantics. Figure 2 depicts the general flowchart of the procedure used in this article, along with the encoder–decoder architecture, the emotion classification module, and the backbone network for extracting picture feature sequences.

3.1 Image sequence generation

Since the encoder anticipates a sequence as input, the abstract painting dataset in this study has been uniformly normalized, meaning that its length and width are 224 and its number of channels is 3. To extract the image's features, the image is supplied into the backbone network. The residual network has a strong feature learning ability and adapts to the characteristics of the backbone convolutional network architecture. In this study, ResNet-50 is adopted as the backbone network to solve the network degradation problem brought by fewer samples of abstract paintings to simplify the model training parameters of this article to a certain extent, improve the training efficiency, and carry out comparative experiments with the residual network variant in the ablation experiments, and to assess the influence of the backbone network on the model accuracy rate (Ahmad et al., 2023). The abstract painting dataset is generated by the backbone network to generate canonical image features with a length and width of 7 and a channel count of 256 and is spread into a one-dimensional sequence, resulting in an image sequence of length 49 and a channel count of 256 to be fed to the encoder.

3.2 Encoders

By adjusting the number of encoder layers, the model demonstrates the significance of global image-level self-attention, guarantees that there is no appreciable loss of accuracy when $N = 6$, and prevents an increase in training difficulty brought on by the addition of too many parameters. This article adopts the position coding method of detection transformer (DETR), which uses the sine and cosine functions to encode the positions of rows and columns of the parity channel of the abstract painting feature map, adapting to the sequence input of the encoder–decoder architecture (Ahmad and Wu, 2023). The encoder–decoder architecture is not sensitive to the order of the image sequence and does not have the ability to learn the sequence position information. The calculation for the position coding as shown in Equation (1):

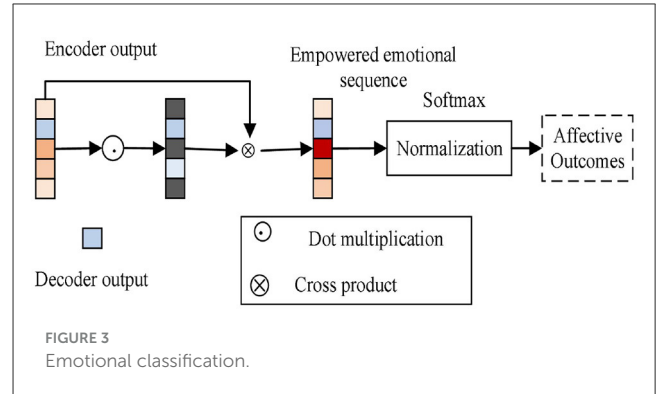
$$f(x)^i = \begin{cases} \sin\left(\frac{x}{10000^{i/128}}\right), & i = 2k (k \in [0, 127]) \\ \cos\left(\frac{x}{10000^{i/28}}\right), & i = 2k + 1 (k \in [0, 127]) \end{cases} \quad (1)$$

where x is the row and column spread value of point (p, q) and i is the channel of the feature map. For a feature map with a length and width of 7 and a channel count of 256, respectively, the row and column position encoding on the point with a channel of 10 and a coordinate value of $(1,2)$ is $\sin [((1 \times 7) + 2)/(1,000,010/128)]$ and $\sin [(2 \times 7) + 1)/(1,000,010/128)]$, respectively, and the position encoding of the remaining image sequences of the channels is computed by this rule. Encoding finally generates a one-dimensional feature sequence with a length of 49 and a channel count of 256 with position information.

The Q, K, V in the encoder is a one-dimensional sequence of a fixed length of 49 and a channel count of 256, which is used as sentiment weights in translating the image sequence and ordering its position in each encoding session. As the model learns the feature dependencies between image sequences, the multi-head self-attention module supports the model by reinforcing the original features with sequence global information. This support enables the model to learn discriminative features for sentiment classification. The original image sequence serves as the input for the first coding layer, and the input for each succeeding layer is the image sequence encoded in the preceding layer. The picture feature sequences are given to the decoder after being encoded and learned by many coding layers of the encoder, avoiding the issue of delayed network convergence and poorer accuracy brought on by the increased depth of the model.

3.3 Decoders

The blank attention in this study has the same format as the feature sequence of the model input, that is, a sequence with a fixed length of 49 and a channel count of 256. Similar to the encoding phase, the blank attention is weighted as a query statement with Q, K, V of the first self-attention layer in the decoder, but at this point, the blank attention does not need to focus on the location information. At each decoding stage, the multi-head attention module transforms the blank attention sequences and generates



the output of the attention sequences with weights by avoiding the problem of slower model convergence through the residuals and normalization module.

The attention sequence with weights from the upper layer and the output sequence from the encoder are fed into the second self-attention layer. This study uses the same sine and cosine functions in the decoder as in the encoder to encode the position of the weighted attention sequences from the upper layers since the output sequence of the encoder contains positional information and needs to accommodate its positional connection. The positional encoding of the picture sequence for each channel is calculated for the weighted attention sequence of length 49 and a channel count of 256. This positional encoding is applied to the rows and columns of the parity channels. Ultimately, a weighted attention sequence of length 49 and a channel count of 256 with position information are obtained. It is combined with the output sequence of the encoder as a query statement and weighted with Q, K, V from the second layer in the decoder. In each decoding stage, the output sequence of the encoder is translated, and the sequence positions are sorted.

3.4 Classification of emotions

Figure 3 depicts the emotion classification module. The sentiment classification module combines the output sequences of the encoder and decoder to produce weighted sentiment sequences, which suppress redundant sentiment information in the model, direct the model to concentrate on deep and shallow sentiment information, and improve the model’s ability to classify sentiment. The fully connected layer is used to map the weighted sentiment sequences, and the cross-entropy loss is minimized to produce stable sentiment classification results (Ahmad et al., 2021). The normalized exponential function is used to calculate the probability value of each type of sentiment; the abstract painting sentiment predicted by the model has the highest probability value.

The normalized exponential function is as shown in Equation (2):

$$S_i = \frac{e^i}{\sum_{j=1}^{12} e^j} \quad (2)$$

where S_i is the normalized value of a particular sentiment and computation $\max(S_i)$ is the abstract painting sentiment label predicted by the model.

One popular loss function for handling classification difficulties is the cross-entropy function, which is primarily used to quantify the difference between two probability distributions. The cross-entropy loss function is as shown in Equation (3):

$$Loss = \frac{1}{batch_size} \sum_i \sum_{c=1}^{12} y_{ic} \log_2(p_{ic}) \quad (3)$$

Each term in the cross-entropy function is p and q and p indicates the true probability distribution and q represents the predicted probability distribution. The cross-entropy function describes the difference between the two probability distributions. For the special case, the cross-entropy function of the binary classification problem, there are a total of two terms, i.e., probability distributions of classes 0 and 1, and there is $p(0) = 1 - p(1)$, so we can get the expression for the binary classification cross entropy loss function, where y_{ic} is the true value and p_{ic} is the probability of the predicted value.

4 Image preprocessing

4.1 Datasets

The abstract dataset, which includes 280 abstract paintings, was created by Machajdik. These paintings are better suited for challenges requiring the prediction of emotion distribution because they simply feature colors and textures and not any clearly discernible objects. The 230 participants in the dataset expressed their emotions by identifying these 280 photographs, with an average of 14 people doing so. The final sentiment category is determined by which of these sentiment markers received the most votes. Due to the ambiguity of emotions, several categories may have extremely similar or identical numbers of votes, making the classification process unclear. Therefore, the ratio of votes for each emotion category is used as a probability distribution to form a probability distribution of emotions corresponding to the image, as shown in Figure 3.

4.2 Feature extraction

Since abstract paintings contain only colors and textures and do not generate emotions through specific objects, the features extracted are emotional features based on the theory of artistry.

4.2.1 Color histogram

Artists use colors to express or trigger different emotions in observers, and extracting color histograms from color features is a common and effective method. The color histogram space H is defined as Equation (4):

$$H = [h(0), h(1), \dots, h(L_k)], \sum_{k=1}^K h(L_k) = 1 \quad (4)$$

where $h(L_k)$ denotes the frequency of the k th color. The similarity of the color histograms of the two images are measured using the Euclidean distance as shown in Equation (5):

$$D(H_s, H_d) = [(H_s - H_d)^T (H_s - H_d)]^{1/2} \quad (5)$$

4.2.2 Itten comparison

Itten successfully used the strategy of color combination by defining seven contrast attributes. Machajdik used seven contrast attributes such as light and dark contrast, saturation contrast, extension contrast, complementary contrast, hue contrast, warm and cool contrast, and simultaneous contrast of images as the emotional characteristics of artistry theory.

As in the case of light and dark contrast, the image is segmented into R_1, R_2, \dots, R_N , small chunks using the watershed segmentation algorithm, and the average h_n (Chroma) b_n (Brightness) s_n (Saturation) is calculated for each chunk. Calculation b_n belongs to five fuzzy luminance: $\left\{ \begin{array}{l} \text{Very Dark (VD), Dark (D), middle (M),} \\ \text{Light (L), Very Light (VL)} \end{array} \right\}$ affiliation function as shown in Equations (6–10).

$$VD = \begin{cases} 1 & b_n \leq 21 \\ \frac{39-b_n}{18} & 21 < b_n \leq 39 \\ 0 & \end{cases} \quad (6)$$

$$D = \begin{cases} \frac{b_n-21}{18} & 21 < b_n \leq 39 \\ \frac{55-b_n}{16} & 39 < b_n \leq 55 \\ 0 & \end{cases} \quad (7)$$

$$M = \begin{cases} \frac{55-b_n}{16} & 39 < b_n \leq 55 \\ \frac{b_n-55}{13} & 55 < b_n \leq 68 \\ 0 & \end{cases} \quad (8)$$

$$L = \begin{cases} \frac{b_n-55}{13} & 55 < b_n \leq 68 \\ \frac{84-b_n}{16} & 68 < b_n \leq 84 \\ 0 & \end{cases} \quad (9)$$

$$VL = \begin{cases} \frac{84-b_n}{16} & 68 < b_n \leq 84 \\ 1 & b_n > 84 \\ 0 & \end{cases} \quad (10)$$

Thus, a 1×5 dimensional vector for each small block of image R_1, R_2, \dots, R_N is obtained, and for the whole image, the light/dark contrast is defined as Equation (11):

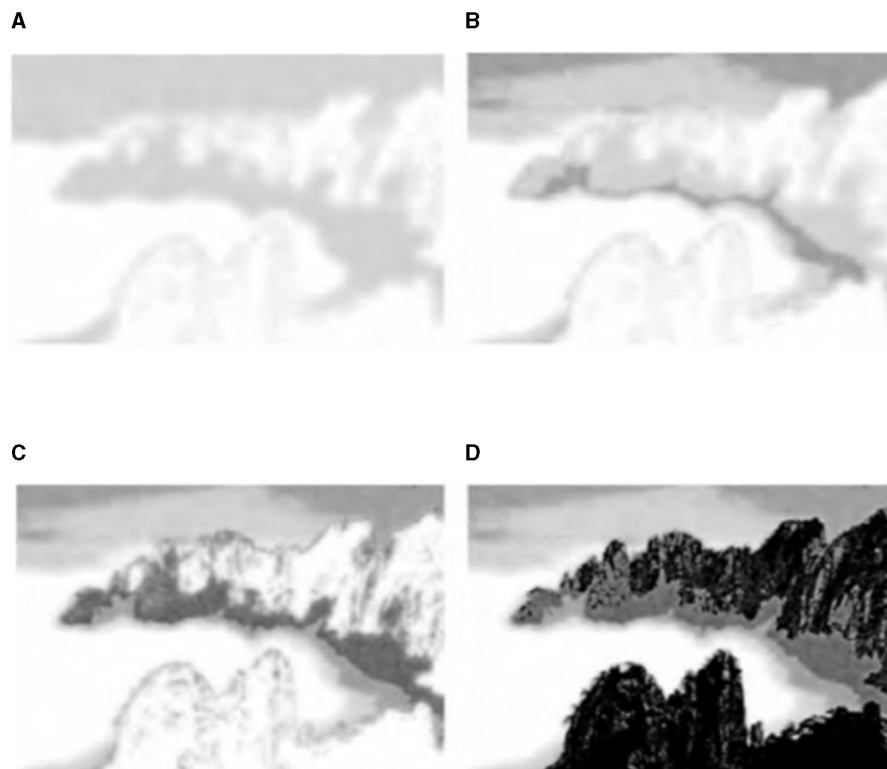


FIGURE 4 Layer stacking process effect and ink simulation results. **(A)** One grayscale layer overlay effect. **(B)** Two grayscale layer overlay effect. **(C)** Three grayscale layer overlay effect. **(D)** Seven grayscale layer overlay effect.

$$B(i) = \left[\frac{1}{\sum_{n=1}^N R_n} \sum_{n=1}^N R_n (B_n(i) - \bar{B}(i))^2 \right]^{1/2} \quad (11)$$

where $i = 1, \dots, 5$, R_n is the number of pixels in the split block.

In this way, the vector expression of the contrasting attributes of the images is obtained as features, and the similarity of the different images is calculated by the Euclidean distance.

The Itten model is also used to determine whether or not an image is harmonic, and it can also be used to identify an image’s emotional expression. Select three to four of the image’s prominent colors, connect them to the colors on the Itten hue wheel, and if they form a positive polygon, the image is harmonic. To determine the dominant chromaticity of an image, make a histogram of its N colors. Ignore the colors with a proportion of $<5\%$. The harmony of a polygon can be assessed by comparing its internal angles to those of a square polygon built from the same number of vertices.

4.2.3 Texture

The main idea behind the statistical approach to texture analysis is to symbolize textures by the randomness of the distribution of gray levels in a graph. We define z as a random variable representing the gray levels, L as the maximum gray level of the image, Z_i as the number of pixels with gray level i , 01 denotes the

gray level histogram, and with respect to z , the n th order moments are calculated as shown in Equation (12):

$$u_n(z) = \sum_{i=0}^L (Z_i - m)^n p(z_i) \quad (12)$$

$$m = \sum_{i=0}^L z_i p(z_i) \text{ is the mean value of } z.$$

The second-order moments are more important in texture description; it is a measure of grayscale contrast, where $R = \frac{1}{1+u_2(z)}$ indicates the smoothness of the image, and a smaller value of $u_n(z)$ corresponds to a smaller R value, indicating that the smaller the value of R , the smoother the image.

4.3 Weighted K -nearest neighbor sentiment distribution prediction algorithm

Assuming that there are M sentiment categories C_1, \dots, C_M and N training images, x_1, \dots, x_N (which also denote the corresponding features of the images) use $p = \{P_{n1}, \dots, P_{nm}, \dots, P_{nM}\}^T$ to denote the sentiment distribution of X_n , where P_{nm} denotes the probability that x_n expresses a sentiment of c_m , and for each image, there is $\sum_{m=1}^M P_{nm} = 1$.

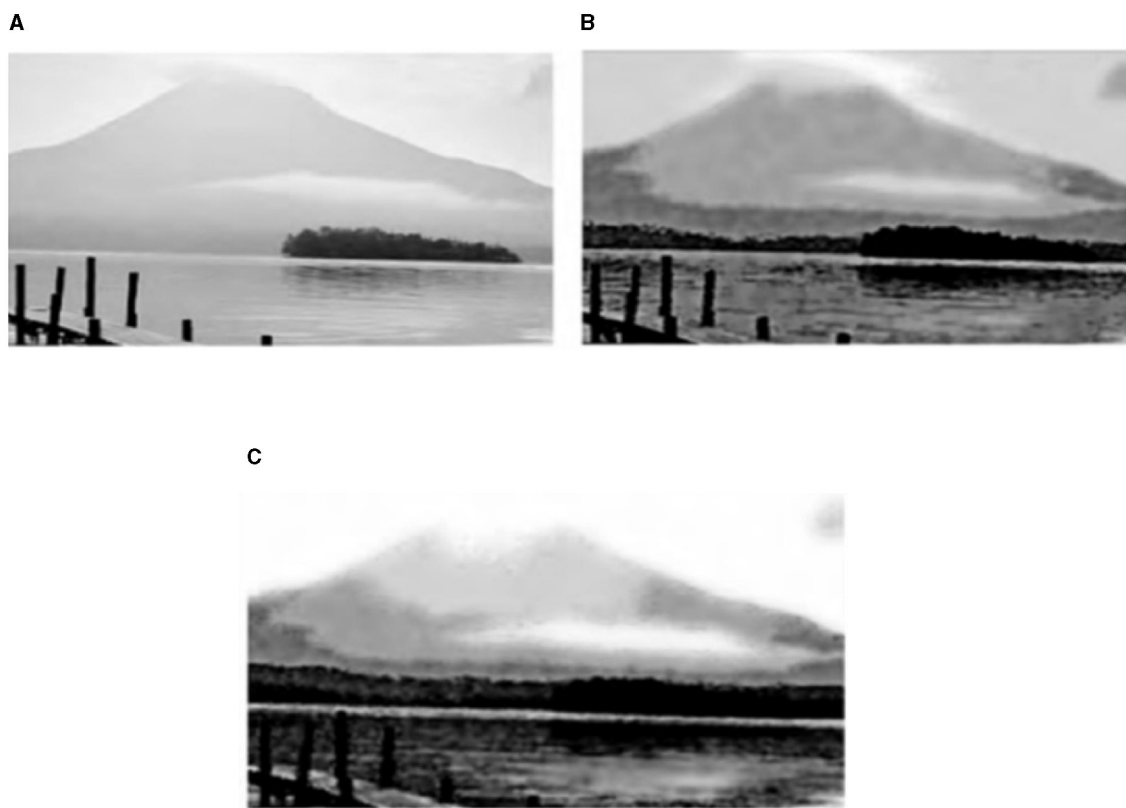


FIGURE 5 Comparison of direct eighth order color reduction and histogram prescribed ink simulation 1. **(A)** Landscape original 2. **(B)** Direct eighth order color reduction link effect. **(C)** Histogram normalization link effect.

Assuming that y is a test image, the goal of this study is to find the sentiment distribution $p = \{p_1, \dots, p_M\}^T$ of y , i.e., as shown in Equation (13).

$$f(\{x_n, p_n\}_{n=1}^N, y) \rightarrow p \tag{13}$$

Training sets that are very far away have little effect on y . Considering that including all training sets can slow down the run and irrelevant training samples can also mislead the algorithm’s classification, the effect of isolated noise samples can be eliminated by taking a weighted average of the K -nearest neighbors.

Weighted K -nearest neighbor option denotes only the drizzle functions corresponding to the K training images that assign the larger weights to the closer nearest neighbors. denotes the sentiment distribution of the K training images nearest to the test image, which is considered as a basis function, and the sentiment distribution P of the test image y is computed by performing a distance-weighted summation of the basis function, i.e.,

$$P = \frac{\sum_{k=1}^K s_k p_k}{\sum_{k=1}^K s_k} \tag{14}$$

where s is the similarity between the test sample and the training sample, as shown in Equation (15).

$$s = e\left(-\frac{d(x_k, y)}{\beta}\right) \tag{15}$$

where d is the Euclidean distance and β is the average distance of y from the training images.

Algorithm: Weighted K -nearest neighbor sentiment distribution prediction algorithm.

Input: Training set (x_n, p_n) , test set y .

Output: Sentiment distribution p for the test set.

1. Calculate the distance d between the test set image y and each image in the training set.
2. Select the first k images $x_1 \dots x_k$ that are closest to y in the increasing order of distance.
3. $\beta = \frac{1}{k} \sqrt{(x_1 - y)^2 + \dots + (x_k - y)^2}$ is brought into Equation (14) in order to compute the similarity s .
4. Calculate the sentiment distribution of the test image y $P = \frac{\sum_{k=1}^K s_k p_k}{\sum_{k=1}^K s_k}$.

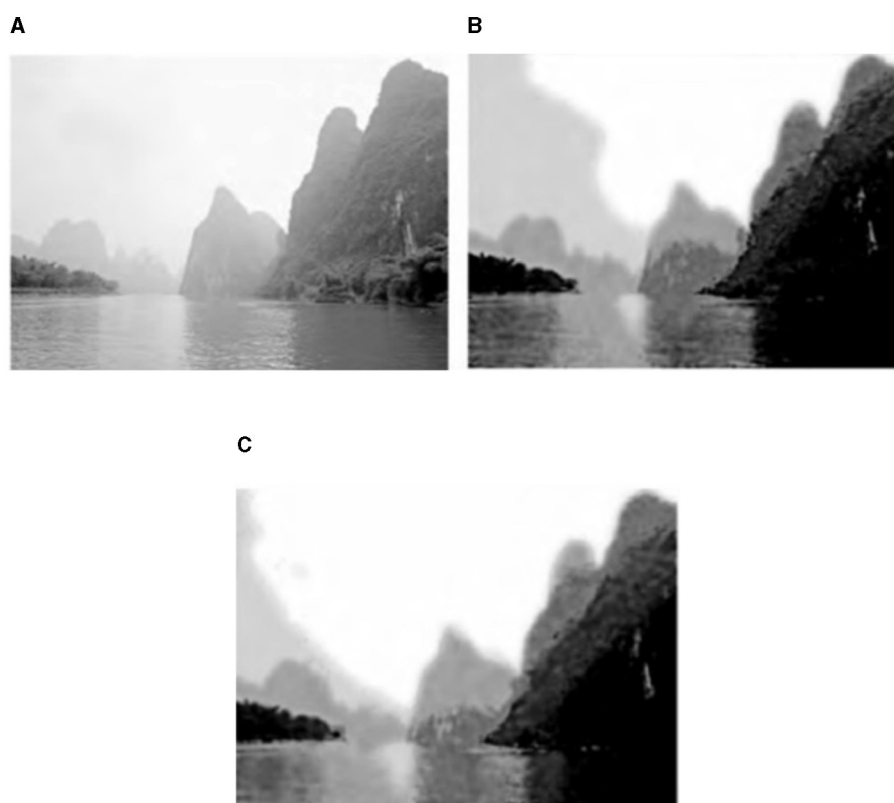


FIGURE 6 Comparison of direct eighth order color reduction and histogram prescribed ink simulation 2. (A) Landscape original. (B) Direct eighth order color reduction link effect. (C) Histogram normalization link effect.

5 Experimentation and analysis

5.1 Landscape image

Experiments on a large number of landscape images (resolution of ~ 1 million pixels, downloaded from “Baidu images”) to achieve the simulation of ink and wash painting and to achieve a more satisfactory simulation effect. Figure 4 represents the algorithm from shallow to deep ink “drawing” simulation process: Figure 4A shows the layer effect, Figures 4B, C represent the first two layers and the first three layers of the superposition effect, and Figure 4D shows the seven-layer superposition effect, that is, the final eight-ink effect [$p_u(u_j) = (0.2, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15$ and Figure 4A for the 00 layer effect. 15, 0.15, 0.15, 0.11, 0.06, 0.03)].

In the algorithm, the histogram specification serves to preset the weight of each ink color and enhance the recognition of the inked area. Figures 5, 6 show a comparison of the ink simulation experiments for two other sets of landscape images and reveal the role of histogram specification in the simulation effect. Figures 5B, 6B show the effect of the algorithm based on direct eighth order grayscale color reduction, and Figures 5C, 6C show the effect of the algorithm based on histogram specification. The values of $p_u(u_j)$ were [0.2, 0.15, 0.15, 0.15, 0.15, 0.15, 0.11, 0.06, 0.03] and [0.3, 0.125, 0.15, 0.125, 0.125, 0.1, 0.05, 0.025]. The direct eight-order color reduction approach is governed by the color values of the original

diagram, which is easily the source of the imbalance of the weight of each ink color and the lack of distinctiveness, as can be seen from the comparison of the two sets of diagrams. The histogram specification method can better control the amount of ink colors and especially strengthen the weight of *Gray* (0) (i.e., white area). Ink simulation has a better sense of hierarchy and differentiation. In summary, the algorithm in this article simulates the ink effect of the landscape map through the method of layer simulation ink overlay, the simulation map has a strong sense of hierarchy, and the layers of ink can be integrated with each other and also has a natural paper-ink penetration effect.

5.2 Abstract paintings

The existing sentiment classification networks ResNet and Swin Transformer and their variants are compared under the sentiment classification accuracy metrics in order to assess the effectiveness of the model in this article. The encoder–decoder structure with various numbers of layers is set up for this article’s method; the one-layer encoder–decoder structure is defined as Tiny and the six-layer encoder–decoder structure is defined as Base. By training five batches of experimental findings and averaging them as the final results of the experimental data, five rounds of cross-validation were used to test the models. To accelerate the convergence of abstract painting sentiment classification, each

TABLE 1 Example of part of the abstract painting dataset.






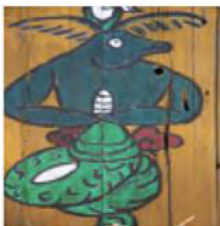


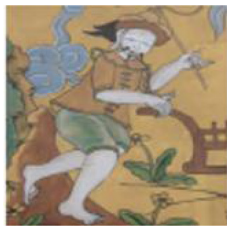
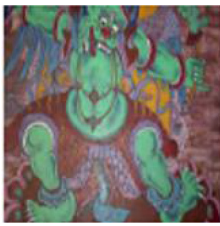
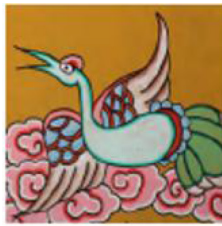
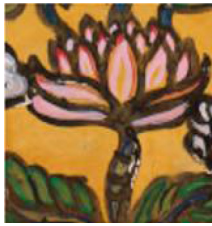
Emotion	Painting theme			
	Character	Ghosts	Animal	Plant
Negative				
	Be cautious	Aversion and resistance	Anxiety and tension	Faint and wilting
Neutral				
	Harmony and friendliness	Neutral and Pure	Be pragmatic and responsive	Impartial and impartial
Positive				
	Diligent and simple	Enthusiastic and proactive	Elegant and gentle	Beautiful and graceful

TABLE 2 Abstract painting emotion classification experiment.

Model	Classification accuracy (%)
ResNet-18	64.6
ResNet-34	68.4
ResNet-50	70.3
ResNet-101	71.4
Swin-T	70.1
Swin-S	72.7
Swin-B	73.2
Vit-T	72.7
Vit-B	76.8
Method of this article—Tiny	74.3
Method of this article—Base	80.8

model is fine-tuned based on the ImageNet pre-trained model, using the Adam W optimizer with a weight decay of 0.1/30 epoch and an initial learning rate of 0.0001, and trained based on the NVIDIA RTX 2080Ti.

TABLE 3 Backbone network experiment.

Method	Backbone	Parameter quantity (M)	Accuracy (%)
1	ResNet-18	29	72.5
2	ResNet-34	39	76.7
3	ResNet-50	42	80.8
4	ResNet-101	61	81.9

The actual Naxi Dongba abstract paintings were gathered from the literature on Na xi abstract paintings, and the abstract paintings were divided into four categories based on the subject matter of the painting's creation. For instance, in the abstract painting data set shown in Table 1, the figures, ghosts and monsters, animals, and plants are shown from left to right, and the abstract paintings were divided into 12 different emotion categories based on the emotions they conveyed.

ResNet50 was used as the backbone network in order to extract image features and tested on the test set for sentiment classification of abstract paintings.

The experimental findings in Table 2 demonstrate that the algorithm presented in this article is superior to ResNet,

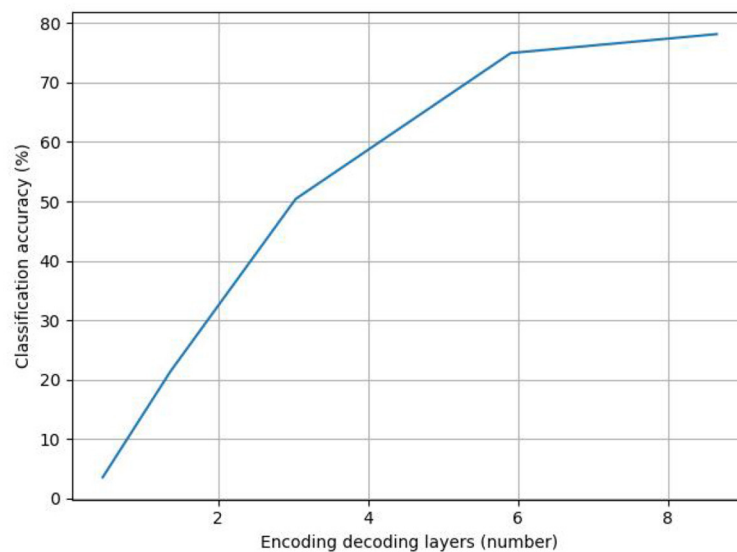


FIGURE 7
Analysis of the number of encoding and decoding layers.

TABLE 4 Ablation experiment.

Method	Decoder output	Encoder output	Accuracy (%)
1	X	✓	74.3
2	✓	X	77.9
3	✓	✓	80.8

Swin, and their variation network topologies for the job of sentiment recognition for abstract paintings. Established sentiment classification techniques like ResNet-101 and Swin-B achieved classification accuracies of 71.4 and 73.2%, respectively, whereas this article's method-Tiny and method-Base produced the best classification outcomes with classification accuracies of 74.3 and 80.8%, respectively.

The existing sentiment classification techniques do not account for the deeper sentiment elements that are buried in abstract paintings; instead, they focus on predicting the sentiment labels of abstract paintings while neglecting their linguistically complex and emotionally varied properties. The method in this article, in contrast, uses blank attention in the decoder and fuses the encoder's output sequence while learning the semantics of the abstract painting image as the emotion attention through the decoder's decoding learning mechanism. As a result, the method employed in this study is able to achieve a higher classification accuracy rate.

This study first conducts ablation experiments on the backbone network, compares a variety of Res Net variants to replace the backbone network, and keeps the structure of this article's model unchanged for the experiments in order to assess the impact of the number of parameters of the backbone network on the accuracy of sentiment classification. The results of the experiments are shown in Table 3.

The model parameter amount was 42 M and the classification accuracy was 80.8% when ResNet-50 was used as the backbone network. The number of model parameters was cut to 29 M with the use of ResNet-18, however the model's classification accuracy dropped by 8.3%. ResNet-34, on the other hand, reduced the number of model parameters by 3 M while increasing the classification accuracy of the model by 4.1% when utilized as the backbone network. The number of model parameters rises by 19 M when ResNet-101 is used as the backbone network, yet the classification accuracy increases by 1.1%. In this article, choosing ResNet-50 as the backbone network ensures that there is no significant decrease in the accuracy rate and avoids the increase in training difficulty due to the introduction of too many parameters.

Figure 7 displays the line graph of the experimental analysis of the number of coding-decoding layers; as the number of coding-decoding layers increases, the model's accuracy gradually increases, suggesting that adding more coding-decoding layers can, to a certain extent, increase the accuracy of the classification of the emotions in abstract paintings. The model uses six coding-decoding layers to achieve 80.8% classification accuracy, avoiding the overfitting issue that results from the stacking of coding-decoding layers. However, as the number of coding-decoding layers increases, the improvement in accuracy eventually slows down and becomes flat.

To prove the effectiveness of this article's attention mechanism for classifying the emotions of abstract paintings, two types of ablation models are set up to eliminate the decoder and encoder outputs, based on keeping the backbone network of the model as ResNet-50: ① The attention mechanism setup is not used in the ablation model, which eliminates the output of the decoder. Instead, the model uses the coded sequence output from the encoder as the basis for emotion classification. The classifier then normalizes the coded sequence to determine the likelihood of outputting emotion labels through the full connectivity layer

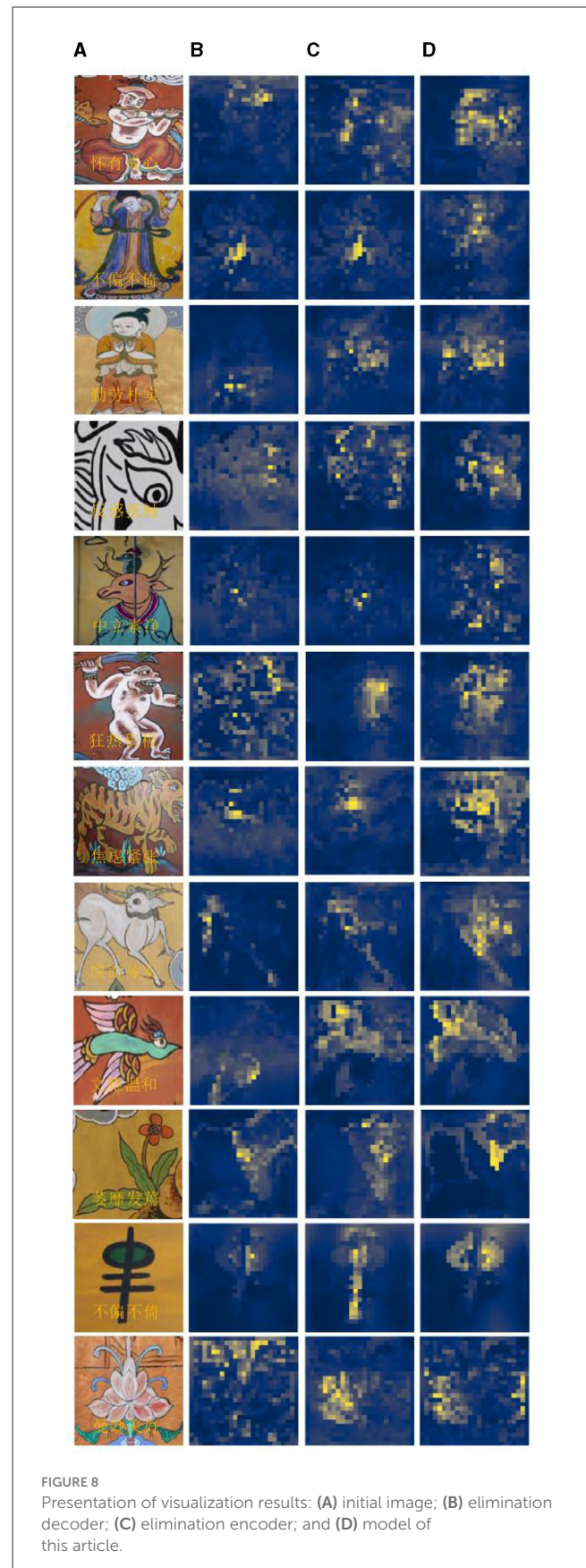
spreading. The fully linked layer disperses the coded sequence, and the normalization of the classifier determines the likelihood of producing emotion labels; ② The attention mechanism established in this research is kept in the ablation model that removes the encoder output, and the model continues to mine the picture semantics of the abstract paintings without fusing the encoder output with the attention mechanism. In Table 4, the experimental findings are displayed.

When the encoder is used to help classify the abstraction of drawing sentiments, the accuracy of sentiment classification decreased after eliminating the decoder output by 6.5%, showing higher classification accuracy than that of the ResNet-50 classification model; however, after eliminating the encoder output, the accuracy of sentiment classification of the ablation model decreased by 2.9%, which is higher than that of the ResNet-101 classification model and close to that of the ResNet-50 classification model. This finding shows that the attention mechanism in this study can help the model recognize abstract paintings' emotions more accurately by acting as a facilitator.

In this study, we used a full convolutional network to calculate the emotional weights of the model, visualize the weight heat map of the model, and simultaneously highlight and locate the regions in the heat map that significantly influence the expression of emotion.

Figure 8A provides an illustration of an abstract painting's original image, which is tagged with the predicted emotions derived from the image by the model test and contains 12 emotions as determined by the experimental data, respectively. The ablation model produced by the elimination decoder is depicted in Figure 8B, with loose regions of attention and unfocused regions of interest in the model's heat map; the regions of interest for abstract paintings of various subjects also differ significantly from one another. Figure 8C demonstrates that, despite being more compact, the model heat map's zone of interest suffers from ambiguous regions of interest and incorrect localization. It is also unresponsive to a smaller percentage of the neutral emotion image. The focus in the figure paintings is on the behavior and movements of the Dong ba figures, and the areas highlighted by the model labeled colors in the different image emotions correspond to the areas of the abstract paintings where the figures are holding arms, dancing, and making gestures, respectively. Figure 8D shows the model heat map of this article, which has a more concentrated region of interest and more stable localization. For emotionally complex animal paintings, the model expands the emotional expression to the animal's body area; for the plant paintings, the color highlighting points out the plant petal area, which corresponds to the plant's budding or blossoming gesture. In the ghost paintings, the model heat map focuses on the ghost behavior and action area.

The visualization experiments demonstrate the comparison experiments of the ablation model and the region of interest of the model described in this article. They also show how the relationship between the abstract painting emotion attention and the image emotion learned by this article model is more intimate and how this has a more immediate effect on the results of the emotion classification. It demonstrates how well the model in this study extracts the emotion from images of abstract paintings, making it more appropriate for classifying the emotions of abstract paintings.



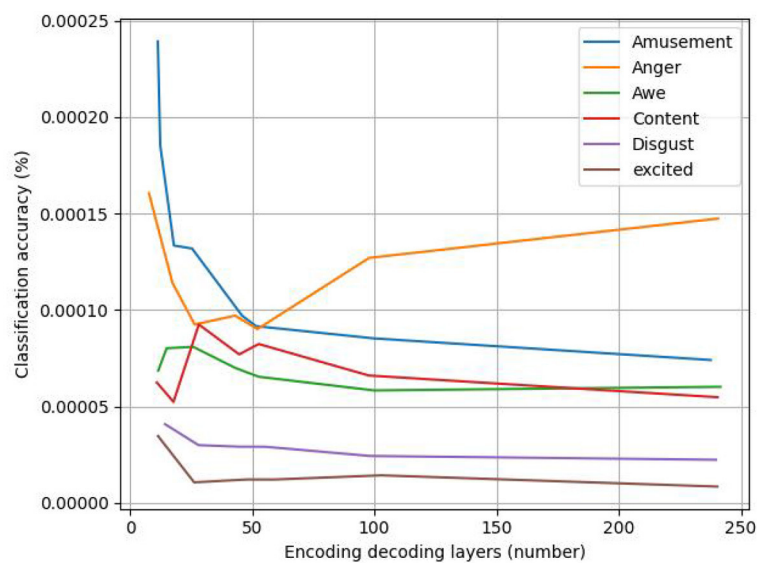


FIGURE 9
Effect of different K -values on the prediction of sentiment distribution.

5.3 Predictive distribution

The effect of different values of K ($K = 5, 10, 20, 40, 50, 100, 252$ using 10-fold cross-validation, where $K = 252$ is the global weighting of the training set) on the prediction of the sentiment distribution in a weighted KKN is shown in Figure 9.

In this example, the optimal K -value is affected by the sentiment category, and considering the average performance, it is considered that the best prediction is achieved at $K = 40, 50$, which outperforms the global weighting, and when $K = 252$, all the training images are used for distribution prediction.

6 Conclusion

The majority of early algorithms employed for sentiment classification were based on shallow machine learning and extract features using manually constructed feature selection techniques that have weak generalization ability, require extensive training times, and entail high labor costs. Because of its superior learning capacity to optimize feature extraction and prevent the flaws of manual feature selection, deep learning has produced positive research outcomes in the field of text sentiment categorization. The attention mechanism's primary objective is to swiftly separate valuable information from a vast amount of data. When applied to the sentiment classification task, it is capable of identifying word dependencies within sentences and identifying the internal organization of the sentence. Using a weighted closest neighbor technique, we provide a novel approach in this study to predict the discrete sentiment distribution of each picture in an abstract painting. Testing shows that the attention mechanism-based classification algorithm achieves a

better classification accuracy of 80.7% when compared to state-of-the-art techniques, thereby resolving the issues of rich material and difficulties in identifying the emotions shown in abstract paintings. Nevertheless, there are several drawbacks to the attention mechanism in this article, such as its incapacity to create the positional link between objects and scenes in abstract paintings. Furthermore, it is restricted by the dataset on abstract paintings and is unable to sufficiently address the issues of imprecise sentiment categorization and imprecise attention learnt from datasets that are made publicly available. Future research methods might thus expand the sentiment dataset to a broader picture data domain and further expand the abstract painting sentiment classification system to a multimodal level in order to overcome these problems.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

KC: Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The author would like to show sincere appreciation to the development of those techniques that have contributed to this research.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships

References

- Ahmad, T., and Wu, J. (2023). SDIGRU: spatial and deep features integration using multilayer gated recurrent unit for human activity recognition. *IEEE Trans. Comput. Soc. Syst.* 2023:3249152. doi: 10.1109/TCSS.2023.3249152
- Ahmad, T., Wu, J., Alwageed, H. S., Khan, F., Khan, J., and Lee, Y. (2023). Human activity recognition based on deep-temporal learning using convolution neural networks features and bidirectional gated recurrent unit with features selection. *IEEE Access* 11, 33148–33159. doi: 10.1109/ACCESS.2023.3263155
- Ahmad, T., Wu, J., Khan, I., Rahim, A., and Khan, A. (2021). Human action recognition in video sequence using logistic regression by features fusion approach based on CNN features. *Int. J. Adv. Comput. Sci. Appl.* 11:121103. doi: 10.14569/IJACSA.2021.0121103
- Alirezazadeh, P., Schirrmann, M., and Stolzenburg, F. (2023). Improving deep learning-based plant disease classification with attention mechanism. *Gesunde Pflanzen* 75, 49–59. doi: 10.1007/s10343-022-00796-y
- Bharadiya, J. (2023). Convolutional neural networks for image classification. *Int. J. Innov. Sci. Res. Technol.* 8, 673–677. doi: 10.5281/zenodo.8020781
- Cetinic, E., and She, J. (2022). Understanding and creating art with AI: review and outlook. *ACM Trans. Multimed. Comput. Commun. Appl.* 18, 1–22. doi: 10.1145/3475799
- Chan, J. Y. L., Bea, K. T., Leow, S. M. H., Phoong, S. W., and Cheng, W. K. (2023). State of the art: a review of sentiment analysis based on sequential transfer learning. *Artif. Intell. Rev.* 56, 749–780. doi: 10.1007/s10462-022-10183-8
- Chandrasekaran, G., Antoanela, N., Andrei, G., Monica, C., and Hemanth, J. (2022). Visual sentiment analysis using deep learning models with social media data. *Appl. Sci.* 12:1030. doi: 10.3390/app12031030
- Chen, X., Li, J., and Hua, Z. (2023). Retinex low-light image enhancement network based on attention mechanism. *Multimed. Tools Appl.* 82, 4235–4255. doi: 10.1007/s11042-022-13411-z
- Ding, F., Yu, K., Gu, Z., Li, X., and Shi, Y. (2021). Perceptual enhancement for autonomous vehicles: restoring visually degraded images for context prediction via adversarial training. *IEEE Trans. Intell. Transport. Syst.* 23, 9430–9441. doi: 10.1109/TITS.2021.3120075
- Li, W., Shao, W., Ji, S., and Cambria, E. (2022). BiERU: bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing* 467, 73–82. doi: 10.1016/j.neucom.2021.09.057
- Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H., et al. (2023). Deep learning attention mechanism in medical image analysis: basics and beyonds. *Int. J. Netw. Dyn. Intell.* 2023, 93–116. doi: 10.53941/ijndi0201006
- Liu, H., Nie, H., Zhang, Z., and Li, Y. F. (2021a). Anisotropic angle distribution learning for head pose estimation and attention understanding in human-computer interaction. *Neurocomputing* 433, 310–322. doi: 10.1016/j.neucom.2020.09.068
- Liu, T., Wang, J., Yang, B., and Wang, X. (2021b). NGDNet: nonuniform Gaussian-label distribution learning for infrared head pose estimation and on-task behavior understanding in the classroom. *Neurocomputing* 436, 210–220. doi: 10.1016/j.neucom.2020.12.090
- McCormack, J., and Lomas, A. (2021). Deep learning of individual aesthetics. *Neural Comput. Appl.* 33, 3–17. doi: 10.1007/s00521-020-05376-7
- Milani, F., and Fraternali, P. (2021). A dataset and a convolutional model for iconography classification in paintings. *J. Comput. Cult. Herit.* 14, 1–18. doi: 10.1145/3458885
- Ngai, W. K., Xie, H., Zou, D., and Chou, K. L. (2022). Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. *Inform. Fusion* 77, 107–117. doi: 10.1016/j.inffus.2021.07.007
- Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., et al. (2022). A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Netw.* 8, 745–762. doi: 10.1016/j.dcan.2021.10.003
- Roy, S. K., Deria, A., Shah, C., Haut, J. M., Du, Q., and Plaza, A. (2023). Spectral-spatial morphological attention transformer for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* 61, 1–15. doi: 10.1109/TGRS.2023.3242346
- Sahoo, K. K., Dutta, I., Ijaz, M. F., Wozniak, M., and Singh, P. K. (2021). TLEFuzzyNet: fuzzy rank-based ensemble of transfer learning models for emotion recognition from human speeches. *IEEE Access* 9, 166518–166530. doi: 10.1109/ACCESS.2021.3135658
- Sampath, V., Maurtua, I., Aguilar Martin, J. J., and Gutierrez, A. (2021). A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J. Big Data* 8, 1–59. doi: 10.1186/s40537-021-00414-0
- Song, T., Zheng, W., Liu, S., Zong, Y., Cui, Z., and Li, Y. (2021). Graph-embedded convolutional neural network for image-based EEG emotion recognition. *IEEE Trans. Emerg. Top. Comput.* 10, 1399–1413. doi: 10.1109/TETC.2021.3087174
- Szubielska, M., Imbir, K., and Szymańska, A. (2021). The influence of the physical context and knowledge of artworks on the aesthetic experience of interactive installations. *Curr. Psychol.* 40, 3702–3715. doi: 10.1007/s12144-019-00322-w
- Teodoro, A. A., Silva, D. H., Rosa, R. L., Saadi, M., Wuttisittikulij, L., Mumtaz, R. A., et al. (2023). A skin cancer classification approach using gan and roi-based attention mechanism. *J. Sign. Process. Syst.* 95, 211–224. doi: 10.1007/s11265-022-01757-4
- Toisoul, A., Kossaiji, J., Bulat, A., Tziropoulos, G., and Pantic, M. (2021). Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nat. Machine Intell.* 3, 42–50. doi: 10.1038/s42256-020-00280-0
- Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., et al. (2022). A systematic review on affective computing: emotion models, databases, and recent advances. *Inform. Fusion* 83, 19–52. doi: 10.1016/j.inffus.2022.03.009
- Yang, H., Wang, L., Xu, Y., and Liu, X. (2023). CovidViT: a novel neural network with self-attention mechanism to detect COVID-19 through X-ray images. *Int. J. Machine Learn. Cybernet.* 14, 973–987. doi: 10.1007/s13042-022-01676-7
- Yang, Z., Baraldi, P., and Zio, E. (2021). A multi-branch deep neural network model for failure prognostics based on multimodal data. *J. Manufact. Syst.* 59, 42–50. doi: 10.1016/j.jmsy.2021.01.007
- Zhang, C., Li, M., and Wu, D. (2023). Federated multidomain learning with graph ensemble autoencoder GMM for emotion recognition. *IEEE Trans. Intell. Transp. Syst.* 24, 7631–7641. doi: 10.1109/TITS.2022.3203800
- Zhao, S., Jia, G., Yang, J., Ding, G., and Keutzer, K. (2021a). Emotion recognition from multiple modalities: fundamentals and methodologies. *IEEE Sign. Process. Mag.* 38, 59–73. doi: 10.1109/MSP.2021.3106895
- Zhao, W., Zhou, D., Qiu, X., and Jiang, W. (2021b). Compare the performance of the models in art classification. *PLoS ONE* 16:e0248414. doi: 10.1371/journal.pone.0248414
- Zhou, J., Pang, L., and Zhang, W. (2023). Underwater image enhancement method by multi-interval histogram equalization. *IEEE J. Ocean. Eng.* 48, 474–488. doi: 10.1109/OJOE.2022.3223733
- Zou, Q., Wang, C., Yang, S., and Chen, B. (2023). A compact pericardial recognition system based on deep learning framework AttenMidNet with the attention mechanism. *Multimed. Tools Appl.* 82, 15837–15857. doi: 10.1007/s11042-022-14017-1

that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.