An improved PageRank algorithm for art appreciation model design

Jingyao Chen Corresp. [1]

[1] The Graduate School of Namseoul University, Cheonan, Republic of Korea

Corresponding Author: Jingyao Chen
Email address: chenjingyao1102@163.com

Aiming at the problem of extracting emotional characteristics in art appreciation, this paper puts forward an innovative method. Firstly, the PageRank algorithm is enhanced using tweet content similarity and time factors; Secondly, the SE-ResNet network design is used to integrate Efficient Channel Attention (ECA) with the residual network structure, and ResNeXt50 is optimized to enhance the extraction of image sentiment features. Finally, the weight coefficients of overall emotions are dynamically adjusted to select a specific emotion incorporation strategy, resulting in effective bimodal fusion. The proposed model demonstrates exceptional performance in predicting sentiment labels, with maximum classification accuracy reaching 88.20%. The accuracy improvement of 21.34% in comparison to traditional Deep Convolutional Neural Networks (DCNN) model attests to the effectiveness of this study. This research enriches the emotion feature extraction capabilities of image and text and improves the accuracy of emotion fusion classification.

An improved PageRank algorithm for art appreciation model design

Jingyao Chen*

The Graduate School of Namseoul University; Cheonan 31020; Republic of Korea;

*Corresponding Author:
Jingyao Chen;
No. 91, Daehangro, Seonghwan-eup, Cheonan
Email address: chenjingyao1102@163.com

Abstract

Aiming at the problem of extracting emotional characteristics in art appreciation, this paper puts forward an innovative method. Firstly, the PageRank algorithm is enhanced using tweet content similarity and time factors; Secondly, the SE-ResNet network design is used to integrate Efficient Channel Attention (ECA) with the residual network structure, and ResNeXt50 is optimized to enhance the extraction of image sentiment features. Finally, the weight coefficients of overall emotions are dynamically adjusted to select a specific emotion incorporation strategy, resulting in effective bimodal fusion. The proposed model demonstrates exceptional performance in predicting sentiment labels, with maximum classification accuracy reaching 88.20%. The accuracy improvement of 21.34% in comparison to traditional Deep Convolutional Neural Networks (DCNN) model attests to the effectiveness of this study. This research enriches the emotion feature extraction capabilities of image and text and improves the accuracy of emotion fusion classification.

Keywords: PageRank; ResNeXt50; ECA; art appreciation; sentiment classification

1. Introduction

Artists use effective integration of their creative techniques and shapes in their artwork to convey the content of their art. This approach not only effectively expresses the emotions of the artwork, but also exceeds the limitations of traditional art language in conveying emotions. Emotions in interactive art encompass not only the emotions of the creator, but also the emotions of the participants involved in the interaction [1-2]. Therefore, artists must not only convey their emotions through artistic techniques, but also possess a thorough understanding of the psychological emotions of the participants. This enables them to effectively convey their emotions to the participants and create an immersive artistic experience.

With the rapid increase in visual information, there is a growing demand for image sentiment processing and analysis. Sentiment analysis plays a crucial role in understanding human emotional experiences with images [3] . In interactive art appreciation, a significant "emotional gap" exists between images and emotions due to the challenge of connecting pixel-level visual information to the complex and high-level mental process of emotions [4]. The ambiguity of interactive art originates from its emotional nature, which has been an area of interest for artificial intelligence research, including image aesthetic quality evaluation [5], stylized image description generation [6] and visual semantic segmentation [7]. Since both emotion and aesthetics are subjective and abstract, methods from both fields can be utilized interchangeably in image aesthetic quality evaluation. In stylized image description generation, sentiment prediction for images can assist in generating image descriptions with sentiment tendencies. Furthermore, in multimedia content filtering and recommendation, users can select specific sentiments for corresponding image search based on their past sentiment preferences, resulting in an improved aesthetic experience for visual participants.

Early scholars, drawing inspiration from the fields of psychology and aesthetics, crafted conventional manual features in order to predict the emotions that images evoke [8-9].Some experts have contended that emotions may be highly interrelated with certain isolated features and have endeavored to establish associations between them, relying on human cognition or associated theories. These investigations typically involve the extraction of image features, such as color, texture, composition, and content, which are then integrated to forecast sentiment [10]. One may depict image emotions by combining generic features and artistic element features at the low level, attribute features and artistic principle features at the middle level, and semantic features and face features at the high level. At present, the primary avenue for art learners to obtain images is through web pages, which yield an enormous amount of visual data and information due to the large number of online community users and fast-paced knowledge dissemination. By analyzing link relationships between web pages and combining them with user search topics, they can offer users more comprehensive and precise information [11]. Furthermore, recognition of different user activity and interaction emotional tendencies can determine the rating and appreciation effects of art images. A novel and efficient representation learning technique for a myriad of image appreciation tasks through convolutional neural networks (CNNs) has been extensively researched [12-14]. This method introduces the PageRank model to examine the original sentiment tendency of the participants' comment vocabulary, aiding in the accurate identification and promotion of their art appreciation behavior.

2. Related works

2.1 PageRank Algorithm

The PageRank algorithm has become a prevalent ranking algorithm that is widely employed in search scenarios where diverse datasets can be represented as graph structures, such as Web search [15], ER graph search [16] , and keyword database search[17]. The personalized PageRank algorithm inherits the principles of the classical PageRank algorithm and employs the data model (graph) link structure to recursively calculate the weight of each node. This algorithm simulates the user's behavior of randomly visiting nodes in the graph by clicking on links, i.e., it follows a random walk model to calculate the probability of random visits to each node in the steady state. The personalized PageRank algorithm accounts for not only the static link structure between nodes when calculating node weights, but also the user preferences expressed in personalized information, such as user queries, favorite pages, and so on [18].

The accuracy of existing PageRank algorithms is determined by the configuration of static parameters, such as the number of Fingerprints and the selection of hub nodes, which cannot be dynamically adjusted at runtime[19]. Additionally, the algorithm's running efficiency is directly determined by the precision requirements set during compilation and cannot be dynamically tuned. Nevertheless, since different users or applications have varying efficiency and accuracy demands for algorithms, estimation algorithms that support run-time tuning of efficiency and accuracy, such as incremental optimization, are imperative [20] . Some scholars have proposed an improvement method based on user interest that involves collecting and analyzing user usage data to determine the direction of user interest, which can enhance the accuracy of recommended content [15]. Another improvement approach is to start from page similarity, which is currently classified into two main categories [21-22]: one uses the space vector model to determine the similarity between the page and the linked page, assigning more weight to the page with greater similarity to solve the problem of average weights; the other is an improved method based on content filtering, which evaluates page text and HTML tags to make the query results more precise.

2.2 Visual sentiment analysis

Image sentiment analysis can be categorized into two approaches: visual features and semantic features. Zhu et al. [23] defined 102 mid-level semantic representations for image sentiment analysis, which resulted in better sentiment prediction results than using visual low-level features alone. Zhao et al. [24] proposed a sentiment analysis approach based on designing visual art sentiment subjects and corresponding visual art sentiment patterns based on different types of sentiment subjects. Other research has focused on learning the affective distribution of visual art affective patterns and further describing visual art characteristics. One feasible method is to calculate the overall sentiment value of an image based on textual sentiment values of adjective-noun pairs and corresponding responses in the image. Jiang et al. [25] constructed a strongly and weakly supervised coupled network system for visual sentiment differentiation of images by importing images into VGGNet and obtaining the entire image features from the fifth convolutional layer, and then using a spatial pooling strategy to obtain weights for each emotion type. Li et al. [26] proposed a 3D CNN combining 3D Inception-ResNet layer and LSTM network to extract spatial features of the image using Inception ResNet and learn temporal relationships using LSTM, and then apply this information for classification.

3. Methodology

Figure 1 depicts the overarching framework of the sentiment analysis model for art appreciation that is utilized in this study. The framework comprises of four distinct components:

(1) Data Importation: The dataset is procured from various microblogs, Twitter, and other media platforms. The crawler tools acquire a substantial amount of text and image data, which is preprocessed into input samples. All the samples are defined as S(Ti, Ii), where Ti and Ii represent all text and images of the ith sample, respectively.

(2) Feature Extraction: The improved PageRank algorithm and ECA+ResNeXt50 is used for text and image sentiment feature extraction, respectively. Text features Ti=(T1,T2,...,Tm) and image features Ii=(I1,I2,...,In) are obtained separately using each component, where m and n denote the dimensions of textual and graphical features, respectively.

(3) Feature Fusion: Firstly, an appropriate fusion strategy is determined based on the statistical sentiment contribution of each modality towards the overall sentiment. Then, cross-modal learning algorithms are employed to fuse the features and compute the statistical sentiment weights between the characteristics of the two modalities.

(4) Sentiment Analysis: Lastly, the trained cross-modal learning algorithm is utilized to achieve sentiment classification of graphical texts.

Figure 1. Art appreciation sentiment analysis model

3.1 Text Emotion Feature Extraction Based on PageRank Algorithm

3.1.1 Algorithm description

Let the seed sentiment word set vector be $S = \{s_1, s_2, \cdots, s_n\}$, whose manually labeled sentiment polarity vector is $Y_S = \{y_1, y_2, \cdots, y_n\}$; the vector of sentiment words to be classified is $W = \{w_{n+1}, w_{n+2}, \cdots, w_{n+m}\}$, the annotation result to be found is $Y_W = \{y_{n+1}, y_{n+2}, \cdots, y_{n+m}\}$. When the sentiment words belong to positive sentiment words. $y_i = 1$; and vice versa. $y_i = -1$. One the one hand, the classification of sentiment words is reliant on the polarity information supplied by the seed sentiment words. On the other hand, it is believed that sentiment words sharing the same polarity are often associated with profound semantic similarity. Thus, the semantic similarity interlinking the sentiment words to be classified can also serve as a valuable determinant of their polarity.

Define the graph $G = \langle N, M \rangle, |N| = |S| + |W|$, where N is the set of G is the set of nodes in the graph (nodes consist of all sentiment words). $|S|$ is the number of seed sentiment words. $|W|$ is the number of sentiment words to be classified. $|W| \times |N|$ Linkage matrix M describes the linkage relationship between nodes in the disjoint graph. M_{ij} is the number of nodes i and j semantic similarity between nodes. M can be decomposed into $|W| \times |S|$ the submatrix of U and $|W| \times |W|$ the submatrix of V. U_{ij} represents the sentiment words to be measured i and the seed sentiment word j semantic similarity between the sentiment word and the seed sentiment word. After introducing the PageRank model, the iterative formula of the sentiment word polarity discriminant algorithm is as follows:

$$Y_W^{(n)} = (1 - \beta) U Y_S + \beta V Y^{(n-1)}_W \quad (1)$$

Where $Y_W^{(n)}$ represents Y_W after n-th iterations. β is the weighting factor. $0 < \beta < 1$.

3.1.2 Improvement based on time factor

To mitigate the issue of lower PR value caused by fewer pages being linked to new content, this paper proposes the inclusion of a time feedback factor in the PageRank calculation formula. The time feedback factor compensates for the PR value of older pages, thereby improving the final recommendation order. The proposed method is based on the fundamental concept that a web page searched multiple times within the same search cycle should be counted only once. The inclusion of the time feedback factor in the PageRank calculation formula is expressed as follows:

$$W_t = e/T \tag{2}$$

Where. e/T is the expression for calculating the time feedback factor of a web page, which indicates the frequency of content searched by search engines. e is usually taken as $0.15/n$,n is the total number of web pages, and the size of e does not affect the distribution of the final PR value, but affects the iterative process of the algorithm, which can effectively improve the situation of low PR value of new pages.

The fundamental concept behind the proposed enhancement, which leverages both similarity and time factors, is to assign PR value based on the degree of similarity between the current web page and the linked pages, rather than a uniform assignment strategy. Additionally, a time feedback factor is incorporated to account for new web pages. The modified PageRank calculation formula is expressed as follows:

$$PR(u) = (1 - d) + d \cdot \left(\sum_{S_v} PR(v) + \alpha \cdot \text{Score}\,(q,D)\right) + W_t \tag{3}$$

3.2 Image emotion feature extraction based on improved ResNet

In this paper, we apply ECA to the residual network structure with the design idea of SE-ResNet network, and improve the optimization of ResNeXt50. The ECA mechanism is shown in Figure 2.

Figure 2. Schematic diagram of ECA attention mechanism

Assuming that any of the feature transformations including convolution is denoted as Ftr. $X \rightarrow U$, where $X \in R^{H \times\times\times c}$,$U \in R^{H \times \| \times c}$, then compress all global information in a channel descriptor using global averaging pooling (GAP), the spatial dimensionality $H \times W$ in shrinking feature U is used to generate statistical variables.

$$Z_n = F_{sq}(u_n) = \frac{1}{H \times W}\sum_i^H\sum_j^W u_n(i,j) \tag{4}$$

Where Z_n can be interpreted as a collection of local features whose statistical information can express the whole image and have a global field of perception. Statistics. The statistic Z does not require dimensionality reduction, and the attention of each channel can be obtained by the following way. The weight size is used as the measure of attention.

$$\rho = \sigma\left(W_k \cdot Z\right) \tag{5}$$

Where W_k contains $k \times C$ parameters, which are defined as .

$$\begin{bmatrix} w_1^1 & \cdots & w_1^k & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w_2^2 & \cdots & w^{k+1}_2 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w^{c-k+1}_c & \cdots & w_c \end{bmatrix} \tag{6}$$

where σ represents the sigmoid nonlinear activation function.

In this paper, 8 ResNeXt50 modules are stacked in series, and the ECA attention mechanism is embedded after each ResNeXt50 module to capture the interdependent associations between channels. It further enhances the image emotion feature extraction capability. The attention weights of each channel after the ECA module are.

$$\rho = \sigma\left(H^{(k)} * F_{sq}(Q)\right) \tag{7}$$

In order to effectively avoid the degradation problem of deep networks, the idea of residuals is introduced to weight each channel and add it to the original input features.

$$Y = (w \otimes Q) \oplus X \tag{8}$$

Where Y denotes the output feature map after a block. $\otimes$ represents the corresponding dot product of the elements. $\oplus$ represents the corresponding sum of elements.

By integrating the ECA mechanism in DCNN, the network can be tailored to meet the specific requirements of different depths. In shallow networks, this module enhances the quality of extracted features at lower layers by highlighting informative features. In contrast, in deeper networks, the significance of this module becomes more pronounced as the extracted features become more strongly correlated with the target category.

3.3 Emotion feature integration

For the integration of emotion features between text and images. First, $f_\theta(X^i)$ is defined as the input features X^i in the parameter θ and then estimate the probability distribution by the Sigmoid function with Equation (9).

$$f_\theta(X^i) = \frac{1}{1 + e^{\theta^T X^i}} \tag{9}$$

Equation (9) defines the degree of bimodal contribution to the overall sentiment. ρ^T and ρ^T are the weights of the text and picture contributions to the overall sentiment, respectively, calculated as.

$$\rho^T = f_\theta(X^T) - \frac{1}{2}$$
$$\rho^I = f_\theta(X^I) - \frac{1}{2} \tag{10}$$

Then the sentiment weight coefficients of the two modalities are compared to determine the appropriate fusion strategy, which is calculated as follows.

$$X^c = \begin{cases} X^T \cup X^I & \text{if } \rho^T * \rho^I > 0 \\ X^T & \text{if } |\rho^T| - |\rho^I| \geq 0 \\ X^I & \text{if } |\rho^T| - |\rho^I| \leq 0 \end{cases} \tag{11}$$

Where X^c denotes the fused features. Sentiment classification is performed by fusing the features of the text and image modalities, provided that the product of their respective sentiment weights and sentiment weight coefficients is positive. If the absolute difference between the sentiment weight coefficient values is greater than zero, the polarity of the sentiment is determined based on the sentiment weight coefficient of the text modality.

The cross-modal learning algorithm calculates the predicted probability distribution of image and text sentiment by Kullback-Leibler divergence. The sentiment probabilities of text and image are discrete events, which are assumed to be defined as Event A and Event B, respectively.

$$D_{KL}(A \parallel B) = \sum_i P_A(x_i) \log\left(\frac{P_A(x_i)}{P_B(x_i)}\right) \tag{12}$$

The loss function of Sigmoid can be used to consider the loss between the expected estimate and the true label, as well as related to the loss between the fusion characteristics of the graphical features and the estimated distribution, as shown in Equation (13).

$$\mathrm{J}(\theta) = \frac{1}{N}\sum_{i=1}^{N} D\left\{Y_i \parallel f_{\theta^c}(X_i^c)\right\} + \frac{\alpha}{2}\theta^T\theta + \frac{\beta}{N}\sum_{i=1}^{N}\left[D\left\{f_{\theta^c}(X_i^c) \parallel f_{\theta^T}(X_i^T)\right\} + D\left\{f_{\theta^c}(X_i^c) \parallel f_{\theta^I}(X_i^I)\right\}\right] \tag{13}$$

Where . $\theta = \{\theta^c, \theta^T, \theta^I\}$ represents the cross-modal learning parameters. α and β are the superparameters of the model. $\frac{\alpha}{2}\theta^T\theta$ is the canonical term, to prevent overfitting of the model.

4 Experiment and analysis

4.1 Experimental setup

For the experimental corpus, a total of 814 positive sentiment words and 1,232 negative sentiment words were selected from the HowNet Sentiment Dictionary. Prior to analysis, the data were preprocessed by separating the words using JIBEA for high-frequency word statistics and mutual information model construction. Additionally, a user code mapping table was constructed to ensure uniform coding across all users.

To facilitate the sentiment classification task, this paper employs the hyperparameter method in ResNeXt network. The optimal kernel scale size for convolution is set at 3×3, while the quantization step is fixed at one. The optimal pooling size for the pooling layer is 2×2, and the quantization step is two. The hyperparameters were determined via tuning of the pre-trained network using ResNeXt. The input images were normalized and have a maximum width of 256×256 pixels. Histogram equalization was used to achieve data enhancement.

To evaluate the effectiveness of the proposed algorithm, a comparative test was conducted against several commonly used neural network models. These include:

(1) DCNN: This model utilizes one CNN to extract sentiment features from both the text and image modalities separately, and predicts the sentiment polarity of each modality. The outputs are then fused using an averaging strategy at the decision level.

(2) CCR: This model uses CNN to extract features from both the image and caption text, and then employs KL scatter to learn the consistent sentiment of both modalities. This model is a feature layer fusion approach.

(3) AttnFusion: This model uses BiLSTM to model video frame sequences and text sequences, and an attention mechanism is used to learn the alignment weights between video frames and text words. The features from both modalities are fused using this attention mechanism to produce a more accurate multi-channel feature representation. The aligned multi-peaked features are then fed into the sequence model for sentiment recognition.

(4) MDREA: This model employs two separate RNNs to encode data from image and text inputs independently. An attention vector is generated by computing the weight parameter between the image encoding vector and the text hidden state. The Softmax function is then applied to the vector to predict the sentiment category.

4.2 Results and Discussion

The model utilizes cross-entropy as the loss function and employs the Adam optimizer as the chosen optimization algorithm throughout the entirety of the training process. Figure 3 presents the training outcomes.

Figure 3. Model training results

The examination reveals that there is negligible disparity between the accuracy of the test set and that of the training set. Similarly, the loss value of the test set aligns with the loss value of the training set, implying that the model is viable for practical applications.

4.2.1 Model comparison

The performance evaluation of the sentiment analysis model described in Section 4.1 is presented in Figure 4 and Figure 5, in comparison with the art appreciation sentiment analysis model introduced in this paper. The experimental assessment was conducted on two datasets obtained from the Flickr and Twitter websites, and the evaluation metrics employed were Precision, Recall, and F1.

Figure 4. Comparison of model performance on the Flickr dataset

Figure 5. Comparison of model performance on Twitter dataset

The proposed model surpasses the other four models in terms of Precision, Recall, and F1, as is evident from the figure. In fact, the Flickr dataset's accuracy improved by a substantial margin of 23.14%, 16.35%, 8.11%, and 4.26%, respectively, when compared to the other four methods. Moreover, the proposed model exhibits superior robustness, even when confronted with smaller Twitter datasets. It outperforms the other four models, thereby establishing the better classification results of the proposed model in the field of cross-modal graphical and textual fusion sentiment analysis. By fusing text features and image features, the model successfully leverages the correlation between different modalities, discovers deeper associations, and performs complementary sentiment information. This feature fusion approach also reduces discrepancies, further proving its feasibility and effectiveness on cross-modal data.

4.2.2 Emotion classification results

Figures 6 and Figure 7 present the confusion matrix results of the bimodal sentiment model. In the Flickr dataset, the model achieves an impressive probability of correct prediction for the four labels, namely "sad," "happy," "angry," and "neutral," with accuracy rates of 73.44%, 70.83%, 88.02%, and 59.38%, respectively. In comparison, the Twitter dataset has a higher prediction accuracy, albeit with a reduced recognition of the "happy" sentiment label. On the whole, the model demonstrates the highest recognition rate for angry labels, while its performance is weakest for neutral labels, which are frequently misidentified as sad labels.

Figure 6. Emotion classification results on Flickr dataset

Figure 7. Emotion classification results on Twitter dataset

The experimental results presented above provide evidence that the model proposed in this chapter performs exceptionally well across all evaluation indices. The proposed model not only achieves a high level of prediction accuracy for each sentiment label but also outperforms both unimodal and other bimodal sentiment analysis models in terms of classification performance and stability. Moreover, the proposed model exhibits remarkable generalization ability, further validating its effectiveness in the field of sentiment analysis.

4.3 Discussion

To summarize, the proposed model in this paper effectively improves the sentiment classification accuracy and achieves the best sentiment recognition ability. Additionally, the model's robustness and generalization ability have been significantly enhanced. Moreover, the addition of temporal feedback factors slightly improves the accuracy of the PageRank algorithm.

This is due to the fact that the temporal factor compensates for the PR value and ensures that new high-quality content is appropriately ranked. The ECA-based attention mechanism has also been successful in improving classification accuracy while reducing the number of model parameters by enabling local cross-channel interaction of sentiment features through one-dimensional convolution. Furthermore, the accuracy of the model in distinguishing between Positive and Negative words differs significantly, which is largely due to the quality of seed word selection. Although the number of seed words with different sentiments can be guaranteed to be equivalent, the quality of the seeds cannot be guaranteed. Nonetheless, the experimental results fully validate the feasibility of the proposed method. It is essential to emphasize that although the connection weights for graph nodes in the algorithm description are calculated using HowNet, other supervised or unsupervised methods can also be used to obtain the connection weights.

In specific application scenarios, teachers can guide students to analyze and appreciate images, not only focusing on the surface content of the images, but also delving deeper into the meaning behind them. Through online interactive comments, teachers can capture the different emotions of individual learners towards the same artwork. With the increasing popularity of social media, many users share their current developments on these platforms, which often reflect their genuine emotions when posting messages. By analyzing such information, we can better understand the interests and preferences of users at a particular period and predict future trends based on user sentiment. This approach can also help in the development of art appreciation software, which includes image loading (PhotoLoader), text loading (TextLoader), and prediction results (ShowResult). The software's structure is shown in Figure 8.

Figure 8. Sequence diagram of art appreciation

The image loading function serves the purpose of loading the image content uploaded by the user into a designated variable, carrying out the necessary pre-processing steps, and finally utilizing the processed output as the input for the prediction result class. Similarly, the text loading function is responsible for loading the user's input text content into a variable, executing the required pre-processing and sentiment score vector transformation processes, and subsequently utilizing the resultant output as an input for the prediction result class. The primary function of the prediction result class is to combine the outputs from the image loading and text loading functions and employ them as inputs to the sentiment analysis algorithm, thereby accomplishing sentiment analysis of the fused graphical data.

5 Conclusion

This study proposes a method for determining the sentiment polarity of words based on the PageRank model, which is employed as an input for text sentiment features in the fusion model. Furthermore, inspired by the ResNeXt model, the main module of the deep sentiment feature extraction network is designed as ResNeXt50, and each residual module is equipped with a BN layer to expedite network convergence. The model's training outcomes on both the Flickr and Twitter datasets indicate that the proposed model considerably improves sentiment classification accuracy and achieves superior sentiment recognition capabilities, thereby enhancing the model's robustness and generalization capacity. In practical applications, the decision layer fusion is achieved by dynamic weight assignment, thereby completing the graphical fusion-based sentiment analysis and facilitating the development and design of the appreciation module in software. However, the paper's training network necessitates a substantial amount of labeled sentiment data, which is currently unavailable on a large-scale and clearly labeled for graphical sentiment analysis. By integrating and analyzing multimodal information such as text, image, and speech-video, it is plausible that the unimodal problem can be resolved more effectively and improve the efficiency of human-computer interaction.

Acknowledgments

I would like to thank my school for supporting my research.

Conflicts of Interest

The author has no conflict of interest.

Funding Statement

This work received no funding.

Data Availability

The dataset employed in this investigation is made readily available and accessible to interested parties.

References

[1] Tan, C. T., & Ferguson, S. (2014). The role of emotions in art evaluation. interactive experience in the digital age: Evaluating new art practice, 139-152.

[2] Yeh, S. L., Lin, Y. S., & Lee, C. C. (2019, May). An interaction-aware attention network for speech emotion recognition in spoken dialogs. in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6685-6689). IEEE.

[3] Rao, T., Li, X., & Xu, M. (2020). Learning multi-level deep representations for image emotion classification. neural processing letters, 51, 2043-2061. Sharma

[4] Sharma, R., Pachori, R. B., & Sircar, P. (2020). Automated emotion recognition based on higher order statistics and deep learning algorithm. Biomedical Signal Processing and Control, 58, 101867.

[5] Li, L., Zhu, H., Zhao, S., Ding, G., & Lin, W. (2020). Personality-assisted multi-task learning for generic and personalized image aesthetics assessment. IEEE Transactions on Image

Processing, 29, 3898 -3910.

[6] Sharma, D., Dhiman, C., & Kumar, D. (2022, December). A Review of Stylized Image Captioning Techniques, Evaluation Parameters, and Datasets. in 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST) (pp. 1-5). IEEE.

[7] Mo, Y., Wu, Y., Yang, X., Liu, F., & Liao, Y. (2022). Review the state-of-the-art technologies of semantic segmentation based on deep learning. Neurocomputing, 493, 626-646.

[8] Jiang, Y. G., Xu, B., & Xue, X. (2014, June). Predicting emotions in user-generated videos. in Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 28, No. 1).

[9] Niu, T., Zhu, S., Pang, L., & El Saddik, A. (2016). Sentiment analysis on multi-view social data. in MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016. Proceedings, Part II 22 (pp. 15-27). Springer International Publishing.

[10] Zhang, W., He, X., & Lu, W. (2019). Exploring discriminative representations for image emotion recognition with CNNs. IEEE Transactions on Multimedia, 22(2), 515-523.

[11] Kumar, A., & Garg, G. (2019). Sentiment analysis of multimodal twitter data. Multimedia Tools and Applications, 78, 24103-24119.

[12] Farkhod, A., Abdusalomov, A. B., Mukhiddinov, M., & Cho, Y. I. (2022). Development of Real-Time Landmark-Based Emotion Recognition CNN for Masked Faces. sensors, 22(22), 8704.

[13] Li, H. (2021). Internet tourism resource retrieval using PageRank search ranking algorithm. Complexity, 2021, 1-11.

[14] Talipu, A., Generosi, A., Mengoni, M., & Giraldi, L. (2019, June). Evaluation of deep convolutional neural network architectures for emotion recognition in the wild. in 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT) (pp. 25-27). IEEE.

[15] Roul, R. K., & Sahoo, J. K. (2021). A novel approach for ranking web documents based on query-optimized personalized pagerank. International Journal of Data Science and Analytics, 11(1) , 37-55.

[16] Chakrabarti, S. (2007, May). Dynamic personalized pagerank in entity-relation graphs. in Proceedings of the 16th international conference on World Wide Web (pp. 571-580).

[17] Li, J., Chen, D., Yu, N., Zhao, Z., & Lv, Z. (2021). Emotion recognition of Chinese paintings at the thirteenth national exhibition of fines arts in China based on advanced affective computing. frontiers in Psychology, 12, 741665.

[18] Scozzafava, F., Maru, M., Brignone, F., Torrisi, G., & Navigli, R. (2020, July). Personalized PageRank with syntagmatic information for multilingual word sense disambiguation. in Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations (pp. 37-46).

[19] Gao, Y., Yu, X., & Zhang, H. (2020). Uncovering overlapping community structure in static and dynamic networks. Knowledge-Based Systems, 201, 106060.

[20] Mo, D., & Luo, S. (2021, October). Agenda: Robust personalized pageranks in evolving graphs. in Proceedings of the 30th ACM International Conference on Information & Knowledge Management (pp. 1315-1324).

[21] Lamurias, A., Ruas, P., & Couto, F. M. (2019). PPR-SSM: personalized PageRank and semantic similarity measures for entity linking. bmc bioinformatics, 20(1), 1-12.

[22] Liu, Q., Zhang, R., Liu, X., Liu, Y., Zhao, Z., & Hu, R. (2019). A novel clustering algorithm based on PageRank and minimax similarity. Neural Computing and Applications, 31, 7769-7780.
[23] Zhu, T., Li, L., Yang, J., Zhao, S., Liu, H., & Qian, J. (2022). Multimodal sentiment analysis with image-text interaction network. IEEE Transactions on Multimedia.
[24] Zhao, Z., Zhu, H., Xue, Z., Liu, Z., Tian, J., Chua, M. C. H., & Liu, M. (2019). An image-text consistency driven multimodal sentiment analysis approach for social media. Information Processing & Management, 56(6), 102097.
[25] Jiang, Y., Li, W., Hossain, M. S., Chen, M., Alelaiwi, A., & Al-Hammadi, M. (2020). A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. information Fusion, 53, 209-221.
[26] Li, L., Qin, S., Lu, Z., Zhang, D., Xu, K., & Hu, Z. (2021). Real-time one-shot learning gesture recognition based on lightweight 3D Inception-ResNet with separable convolutions. Pattern Analysis and Applications, 24(3), 1173-1192.

Figure 1

Art appreciation sentiment analysis model

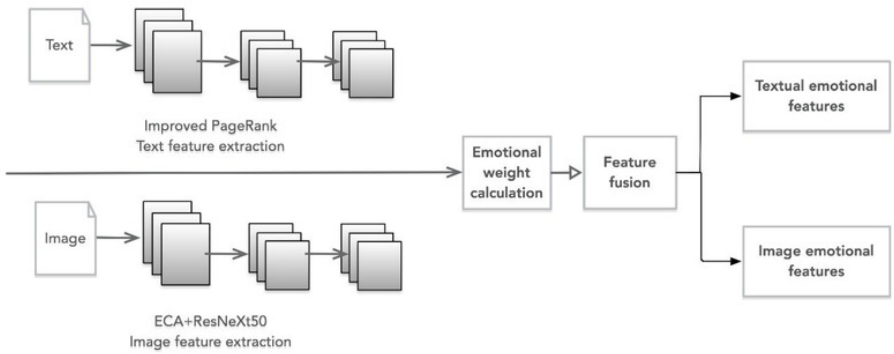

Text
Improved PageRank
Text feature extraction
Image
ECA+ResNeXt50
Image feature extraction
Emotional weight calculation
Feature fusion
Textual emotional features
Image emotional features

Figure 2

Schematic diagram of ECA attention mechanism

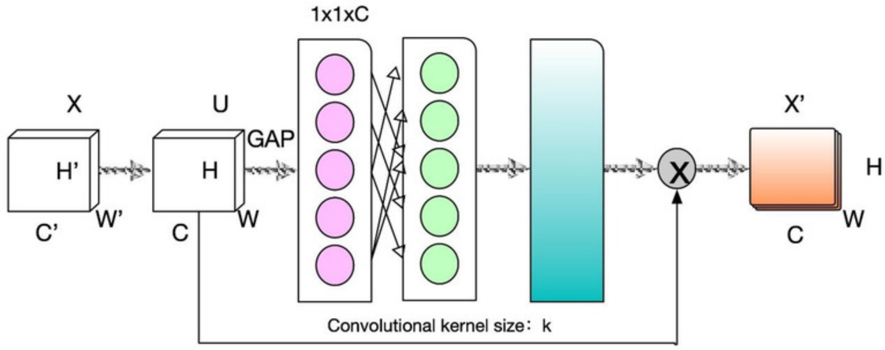

1x1xC
X
H'
W'
C'
U
H
W
C
GAP
X'
H
W
C
Convolutional kernel size：k

Figure 3

Model training results

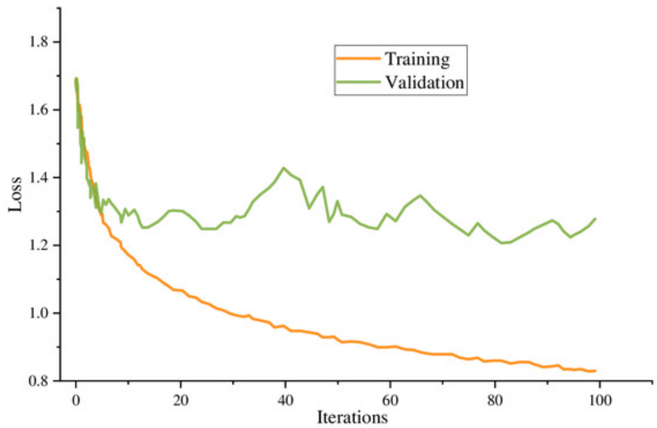
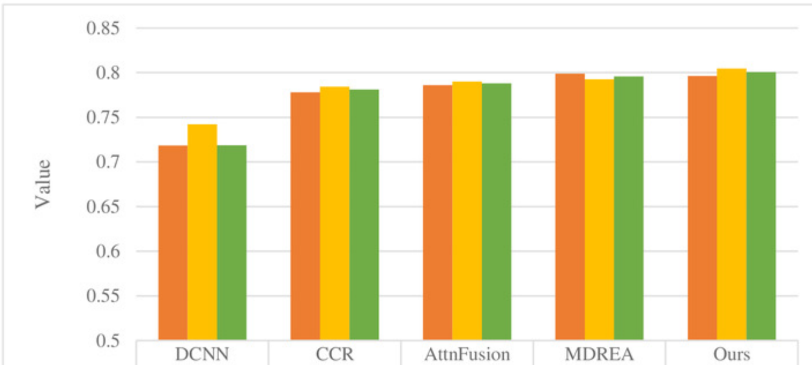

Training
Validation
Loss
Iterations
1.8
1.6
1.4
1.2
1.0
0.8
0
20
40
60
80
100

Figure 4

Comparison of model performance on the Flickr dataset



	DCNN	CCR	AttnFusion	MDREA	Ours
Precision	0.7185	0.7781	0.7859	0.7989	0.7963
Recall	0.7421	0.7842	0.79	0.7927	0.8047
F1	0.7188	0.7811	0.7879	0.7958	0.8005

Figure 5

Comparison of model performance on Twitter dataset

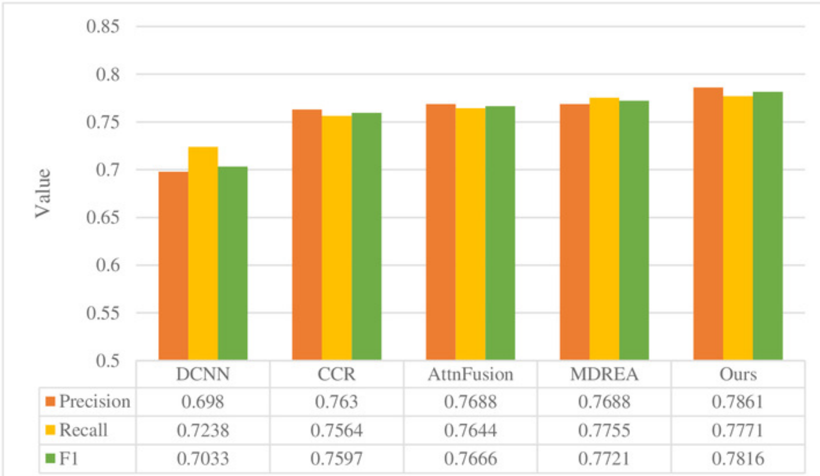

0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Value
DCNN
CCR
AttnFusion
MDREA
Ours
Precision 0.698 0.763 0.7688 0.7688 0.7861
Recall 0.7238 0.7564 0.7644 0.7755 0.7771
F1 0.7033 0.7597 0.7666 0.7721 0.7816

Figure 6

Emotion classification results on Flickr dataset

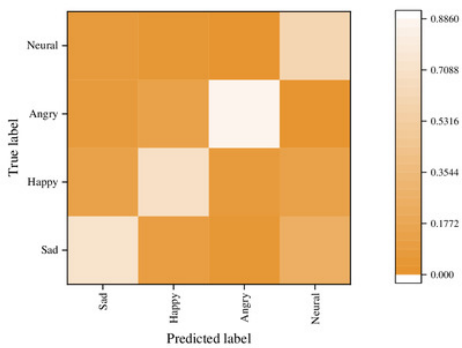

Neural
Angry
Happy
Sad
True label
Sad
Happy
Angry
Neural
Predicted label
0.8860
0.7088
0.5316
0.3544
0.1772
0.000

Figure 7

Emotion classification results on Twitter dataset

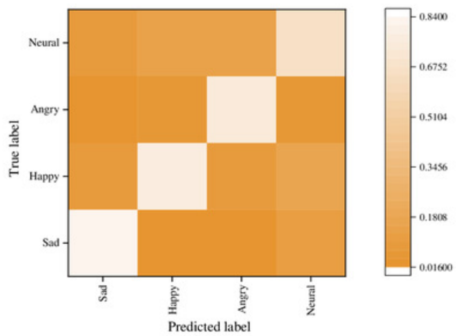
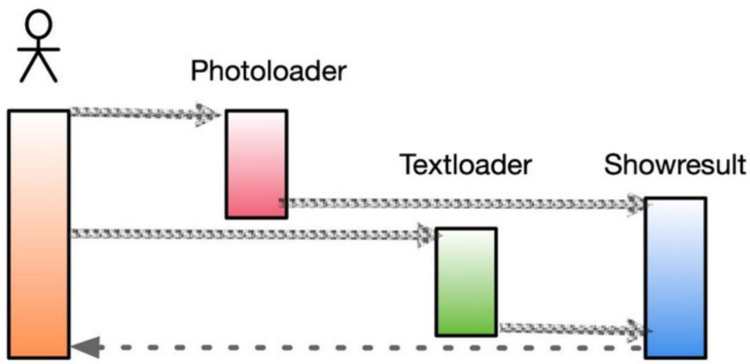

Neural
Angry
Happy
Sad
True label
Sad
Happy
Angry
Neural
Predicted label
0.8400
0.6752
0.5104
0.3456
0.1808
0.01600

Figure 8

Sequence diagram of art appreciation


Photoloader
Textloader
Showresult