Sentiment Analysis of Web Images by Integrating Machine Learning and Associative Reasoning Ideas

Yuan Fang[1]*, Yi Wang[2]
The College of Art and Media-Xianda College of Economics and Humanities, Shanghai International Studies University, Shanghai, 200000, China[1]
EMC Information Technology, R&D Shanghai Co. Ltd, Shanghai, 200000, China[2]

***Abstract*—To achieve automatic recognition and understanding of image sentiment analysis, the study proposes an image sentiment prediction network based on multi-excitation fusion. This network simultaneously handles multiple excitations, such as color, object, and face, and is designed to predict the sentiment associated with an image. A visual emotion inference network based on scene-object association is proposed using the association reasoning method to describe the emotional associations between different objects. The multi-excitation fusion image sentiment prediction network achieved the highest accuracy of 75.6% when the loss weight was 1.0. The network had the highest accuracy of 76.5% when the object frame data was 10. The average accuracy of the visual sentiment inference network based on scene-object association was 91.8%, which was an improvement of about 3.7% compared to the image sentiment association analysis model. The outcomes revealed that the multi-stimulus fusion method performed better in the image emotion prediction task. The visual emotion inference network based on scene-object association can recognize objects and scenes in images more accurately, and both the scene-based attention mechanism and the masking operation can improve the network performance. This research provides a more effective approach to the field of image sentiment analysis and helps to improve the computer's ability to recognize and understand emotional expressions.**

***Keywords*—Sentiment analysis; multi-excitation fusion; image emotion prediction; associative reasoning; attention mechanism**

I. Introduction

Due to the quick advancement of computer technology, the Internet is now a necessary component of daily life. People can watch videos, browse pictures online, and even communicate with friends thousands of miles away in real time [1]. However, with the popularization of the Internet, many problems have emerged. Online images play an important role in human life, but they also have some negative impacts, such as the pornification of online images, violence and false information [2-3]. In order to solve these problems, sentiment analysis of web images is needed, i.e., computer algorithms are used to determine the emotional attributes of web images. Sentiment representation is a key concept in image sentiment analysis, which refers to the use of a certain way to represent and describe the emotions or feelings expressed in an image, so that computers can understand and process the emotional information of the image. Common sentiment representation methods include extracting sentiment words from textual descriptions of images (such as titles, descriptions, labels, etc.) and calculating the sentiment polarity of the words using sentiment dictionaries (such as SentiWordNet). Encode emotional information using low-level visual features of images, such as local areas, colors, textures, etc., and learn a vectorized emotional representation by integrating visual and textual features of the image with conceptual features in the emotional ontology [4]. However, traditional sentiment analysis methods often rely on hand-designed feature extraction, such as color, shape, and texture, etc., which often fail to comprehensively reflect the emotional attributes of web images. The topic of web image emotion (IE) analysis has witnessed significant advancements in machine learning (ML) methods in recent times. ML-based sentiment analysis uses a large amount of text data with labeled sentiment to train models to automatically analyze the sentiment tendencies in the text, thus helping people to better understand and process large-scale sentiment information [5].

Halim et al. proposed a framework for recognizing sentiment in short texts using email text, employing an ML approach and six sentiment classifications. Experimental results indicated that the framework had better performance in sentiment recognition with an average accuracy of 83% [6]. Britzolakis' group presented a tool for sentiment analysis based on lexicon and ML algorithms and explored alternative implementations and open topics for political sentiment analysis on Twitter. The results demonstrated that the methodology could help readers to understand the field and identify the best options to conduct related work and research [7]. Alasmari et al. utilized ML methods for sentiment analysis of Arabic tweets from Saudi Arabian tourism industry using decision trees, random forests, logistic regression, and Naïve Bayes for three categories of classification. The results showed that logistic regression and Naive Bayes performed the best when dealing with Arabic morphology, achieving 86% accuracy [8]. Chirgaiya et al. used natural language processing techniques to train a classifier model for sentiment classification of movie reviews through feature extraction and ranking. The method's 97.68% accuracy in sentiment classification, as demonstrated by the testing data, was a supplement to the web's current movie rating systems [9]. Using a dataset of Facebook user book reviews and taking demographic data into account, Kumar's group investigated the effects of age and gender on sentiment analysis. Utilizing ML techniques for sentiment analysis, the study produced fresh findings about the effects of sentiment analysis on age and gender [10].

Associative inference refers to finding relationships and correlations between various variables in data, and feature

extraction is the selection and transformation of important features from raw data for subsequent analysis. Liewlom's group proposed a new inference framework for determining association rules without defining values for measure, minimum support, and minimum confidence. The study validated the feasibility and effectiveness of the approach through a tree of association rules found from a cancer dataset that reflected the sequential relationships of 15 items associated with the dataset [11]. Li et al. suggested a pedestrian recognition framework based on attribute mining and inference, which improved the performance of pedestrian re-recognition by designing a spatial channel attention module and utilizing the semantic inference and message passing functions of graph convolutional networks. Experimental results indicated that the method achieved 87.03% accuracy on the Market-1501 dataset [12]. The wear monitoring system that Lin' et al. proposed a rapid Fourier transform to extract features from vibration and auditory signals by using a variety of sensors and feature fusion techniques. Through the use of cross-validation, the system developed with a hierarchical neural network structure and sensor feature fusion demonstrated its efficacy and performance under various tightening torque values and spindle speeds [13]. To achieve automatic clustering processing of grid intrusion features, Zhang et al. proposed a fuzzy c-mean clustering based method for extracting intrusion features. This was combined with a fuzzy association rule scheduling method to reconstruct the structure of intrusion statistical feature sequences, and a global optimization method to achieve automatic clustering processing. The approach could enhance the power grid's resistance to attacks and enable autonomous clustering of intrusion feature extraction, according to the results [14]. Guo et al. suggested a new method to extract feature points based on topological information, which achieves feature point detection of point neighborhood topology by introducing an improved local binary pattern to process the original point cloud. Experiments demonstrated that the method performed robustly in extracting point cloud shape features [15].

In summary, many researchers have conducted various designs and studies for sentiment analysis and associative feature extraction. However, these methods and models still have limitations. For example, emotional representation methods are relatively single and rely on manually designed feature extraction, making it difficult to fully reflect the emotional attributes of images. There are many applications in sentiment analysis, and some algorithms need to improve their accuracy when dealing with complex emotions and cross-cultural scenarios. In addition, sentiment analysis research in different fields and languages is relatively scattered, lacking universality and comparability. To improve the effectiveness of sentiment analysis techniques, the study proposes an approach to Web IE analysis that combines ML and associative reasoning concepts.

The study is divided into five sections, with Section II proposing sentiment analysis methods. Section III is the validation and application analysis of the method. Section III is a discussion of research methods, application analysis, etc. Finally, Section V concludes the paper.

II. Methods and Materials

A. IEPN-MIF-Based Method

The Internet is growing quickly, and as a result, a lot of photos and videos with rich emotional content are shared online. Aiming at the problem of image sentiment analysis, the study proposes an IE prediction network based on multi-excitation perception (IEPN-MIF), as shown in Fig. 1. The network consists of three stages, including incentive selection, feature extraction, and emotion prediction. In the incentive selection stage, using object detection and face detection methods in deep learning, specific emotional incentives are accurately selected, which cover aspects such as color I_g, objects I_s, and faces I_e. Then, the feature extraction stage is entered, which synchronously extracts different emotional features from different stimuli. Finally, in the emotion prediction stage, the hierarchical cross-entropy (CE) loss function is used to optimize the emotion prediction results, while fully combining the inherent hierarchical structure of emotions to distinguish between simple negative samples and difficult negative samples. The research also innovatively proposes a new hierarchical CE loss function to further optimize the performance of the whole network and achieve more accurate and efficient IE prediction based on multi-stimulus perception [16].

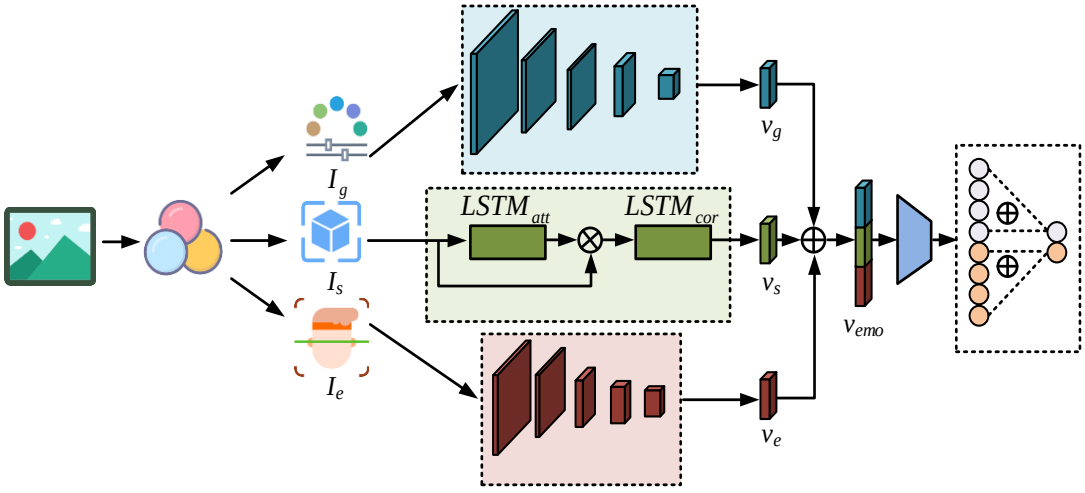

Fig. 1. Image emotion prediction network based on multi-incentive fusion.

In emotion feature extraction, the study extracts color features and other global features in images by constructing a global network. To balance computational efficiency and accuracy, the global network makes use of the ResNet-50 network, which has five convolutional layers and a global average pooling layer [17]. The features output from the last convolutional layer are shown in Eq. (1).

$$F_g = FCN_{res}(I_g) \tag{1}$$

In Eq. (1), the final convolutional layer outputs features as F_g and the full convolutional network as FCN_{res}. The final extracted global features are shown in Eq. (2).

$$G_g = G_{avg}(F_g) \tag{2}$$

In Eq. (2), the global feature is G_g and the fully convolved global average pooling layer is G_{avg}. To better mine the correlation between different objects, the object features are considered as a sequence data and a long short-term memory (LSTM) has been used to carve out this dependency [18]. The study also designed an emotion-specific semantic network for mining semantic associations between different objects and thus inferring IEs. The network consists of two LSTM layers, i.e., attention and association LSTM. The computational procedure of Attention LSTM is shown in Eq. (3).

$$h_t^{att} = LSTM_{att}(x_t^{att}, h_{t-1}^{att}) \tag{3}$$

In Eq. (3), the output of the attention LSTM is h_t^{att} and the input vector is x_t^{att}. The output of the attention LSTM at time t is calculated by the attention LSTM module based on the input vector at time t and the output of the previous time (t-1). The attention output at the current time depends on the current input and the attention state of the previous time. The weight $a_{i,t}$ of the attention module is calculated as shown in Eq. (4).

$$a_{i,t} = w_a \tanh(W_i f_i + W_h h_t^{att}) \tag{4}$$

In Eq. (4), the module learnable parameters are w_a, W_i, and W_h, respectively, and the object features are f_i. The weighted object features are calculated as shown in Eq. (5).

$$f_{att} = \sum_{i=1}^{N} a_{i,t} f_i \tag{5}$$

The weighted object feature f_{att} in Eq. (5) is sent to the associative LSTM. The output vector h_t^{cor} in the associative LSTM is computed as shown in Eq. (6).

$$h_t^{cor} = LSTM_{cor}(x_t^{cor}, h_{t-1}^{cor}) \tag{6}$$

In Eq. (6), the association LSTM input vector is x_t^{cor}, and the output vector of the previous moment is h_{t-1}^{cor}. By feeding object excitations into the attention LSTM and association LSTM separately, the possible redundancy of information between multiple objects can be reduced and the semantic correlations between different objects can be mined [19]. Therefore, the output of the semantic network can be regarded as the higher-level semantic features of the object excitations, as shown in Eq. (7).

$$G_s = h_T^{cor} \tag{7}$$

In Eq. (7), the last moment of LSTM is T, and the semantic features extracted from the object excitation are G_s. In order to extract the face expression features, the study adopts ResNet-18 as the base network to construct the expression network, and the expression features are calculated G_e if shown in Eq. (8).

$$G_e = \begin{cases} G_{avg1}(FCN_{res1}(I_e)), \exists I_e \\ 0, \quad others \end{cases} \tag{8}$$

In Eq. (8), the fully connected layer (FCL) of the expression network is FCN_{res1}, and the average pooling layer is G_{avg1}. Under the given expression category, the expression features are calculated by the ResNet-18 network layer, while in other cases the feature values are 0. The study analyzes three typical emotion incentives (color, object, and face), and designs a dedicated network to extract their emotion features (global features, semantic features, and expression features). These features are independent and complementary to each other, and together determine the final emotion classification result. In order to perform sentiment prediction, these several sentiment features will be spliced, and the final generated sentiment features are shown in Eq. (9).

$$G_{emo} = Concate[G_g, G_s, G_e] \tag{9}$$

In Eq. (9), the splicing operation is $Concate$. The final generated sentiment feature G_{emo} will be input into the subsequent sentiment classifier. In traditional categorization tasks, CE loss is widely used and has made breakthroughs in several tasks. However, sentiment categories are not completely independent from each other, but there is an inherent hierarchical relationship. An eight-category psychological model—which comprises the emotion categories of enthusiasm, amazement, contentment, anger, disgust, sadness, and fear—will serve as the foundation for the dataset that the study will construct. Positive and negative emotions are the emotional polarities of the emotion categories. In the eight-category emotion model, the first four emotions are positive emotions and the last four belong to negative emotions. The eight-category emotion model is shown in Fig. 2.

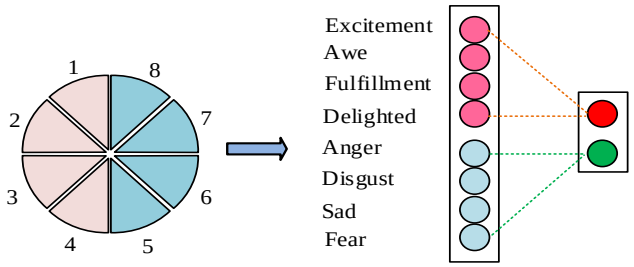

Fig. 2. Eight classification of the emotion model.

After cascading the sentiment features, they are sent to the classifier and activation function respectively to perform operations to obtain the sentiment vector as shown in Eq. (10).

$$p_{emo}(i|G_{emo},W)=\frac{\exp(w_i G_{emo})}{\sum_{i=1}^{C}\exp(w_i G_{emo})} \quad (10)$$

In Eq. (10), the sentiment vector is p_{emo}, and the sentiment type is C. The learnable parameter in the sentiment classifier is W, and its constituent elements are w_i. The traditional CE loss cannot distinguish between positive samples and negative samples, so this study proposes the auxiliary polarity loss and the hierarchical CE loss to solve this problem. Among them, the auxiliary polarity loss is used to distinguish between simple negative samples and difficult negative samples, while the stratified CE loss combines sentiment loss and polarity loss to obtain more accurate sentiment prediction. The final obtained stratified CE loss L_{CE} is shown in Eq. (11).

$$L_{CE}=L_{emo}+\lambda L_{pol} \quad (11)$$

In Eq. (11), the affective loss is L_{emo}, the polarity loss is L_{pol}, and the equilibrium setting hyperparameter is λ.

B. *Image Emotion Classification Based on Associative Reasoning*

When studying IE, one must focus on the scenes and objects it contains and use reasoning to understand IE in the context of how the two interact. The study suggests a scene object-interrupted visual emotion reasoning network (SOLVER) based on this, as seen in Fig. 3.

In the operating mechanism of the SOLVER network, it first uses a faster region convolutional neural network (Faster R-CNN) object detector to extract semantic concepts and visual features of various objects, and then filters and transforms these extracted features. On this basis, sentiment maps are constructed based on object features to represent emotional associations between different objects. Then, a graph neural network (GNN) is used for inference, connecting different object nodes by sentiment edges to generate object features enriched with emotions [20]. In addition, the study also designed a scene-object fusion module, which utilizes the attention mechanism (AM) to fuse object features based on scene features, while constructing the mutual relationship between the scene and objects. In the specific operation, the Faster R-CNN object detector is used to select a set of target candidate regions, pre-trained on the dataset, and finally output attribute categories. Through these steps, the SOLVER network is able to complete its complex operations and feature extraction process. As illustrated in Fig. 4, the sentiment graph is constructed by outputting the ten objects with the highest confidence ranking as its nodes. Each sentiment image is represented by a sequence of object semantic concepts, corresponding confidence scores, and VFs following Faster R-CNN.

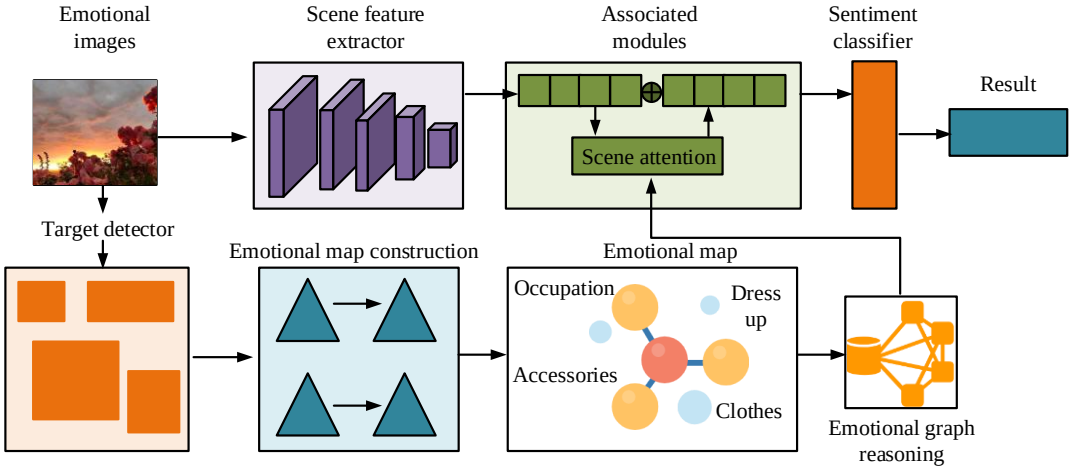

Fig. 3. SOLVER network.

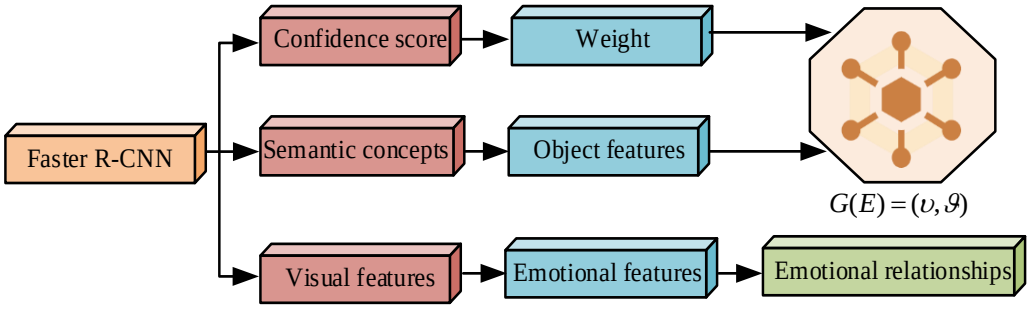

Fig. 4. The construction of emotion map.

The edges of the emotion graph are constructed by VFs and the nodes of the emotion graph are constructed by semantic features, and the emotion VF vector v_i^e is constructed as shown in Eq. (12).

$$v_i^e = \ell_1(W_e v_i + b_e) \tag{12}$$

In Eq. (12), the learnable embedding matrix is W_e, the embedding bias is b_e, the nonlinear regular function is ℓ_1, and the object VFs are v_i. The emotional relationship between different objects is shown in Eq. (13).

$$r_{i,j}^e = \phi(v_i^e)^T \cdot \varphi(v_j^e) \tag{13}$$

In Eq. (13), the sentiment relation of two objects in the same picture is $r_{i,j}^e$, and the two sets of embedding functions are ϕ and φ, respectively. The composition of the sentiment map vector is shown in Eq. (14).

$$G(E) = (\upsilon, \vartheta) \tag{14}$$

In Eq. (14), the emotion graph is $G(E)$, the objects containing different semantic features are υ, and the emotion relationship between different objects is ϑ and is described by an affinity matrix. Then, nodes with a confidence level lower than 0.3 are filtered out by setting a threshold for the confidence level, and the edges connected to the redundant nodes are subjected to a masking operation to reduce the information redundancy of the nodes. Finally, the masked affinity matrix is obtained by weighting the masked matrix to the affinity matrix to describe the sentiment relationship between different objects. Subsequently, the GCN will reason about the sentiment graph to exchange and disseminate information through the structured graph structure. In the design of GCN, residual structures are incorporated to better maintain the original node characteristics. Object features are propagated over the whole sentiment graph based on sentiment relationships in multilayer GCN, where each node is updated based on itself and its neighbors [21]. The multilayer GCN structure iterates the object features based on the sentiment relations to realize the inference of the sentiment graph. The result of emotion graph inference is shown in Eq. (15).

$$O'^{(l)} = f(O'^{(l-1)}, R^{e'}) \tag{15}$$

In Eq. (15), the relationship function between different nodes is f, the output of the last GCN layer is $O'^{(l)}$ and the input edge features are $R^{e'}$. The emotion-enhanced object features are formed by modeling different emotional relationships between objects. Scene is regarded as an additional motivator in the emotional arousal process, which greatly affects the image's emotional tone. When doing an IE analysis, scene features shouldn't be overlooked [22]. Consequently, as illustrated in Fig. 5, the study suggests a scene-based AM. Using ResNet-50 as a base network to build a scene network, this technique mines the emotional connections between objects and scenes, taking an emotional image as input and producing the scene attributes of that image.

Based on scene features and emotion-enhanced object features, AM first projects scene features onto the same embedding space to reduce the difference between scene features and object features. Then the attention weight of each object feature is computed by emotion association to fuse the scene features and object features to establish deep emotional relationships. The attentional weights are calculated as shown in Eq. (16).

$$a_i = \sigma(F_s(f_{sce}) \cdot F_o(o_i')) \tag{16}$$

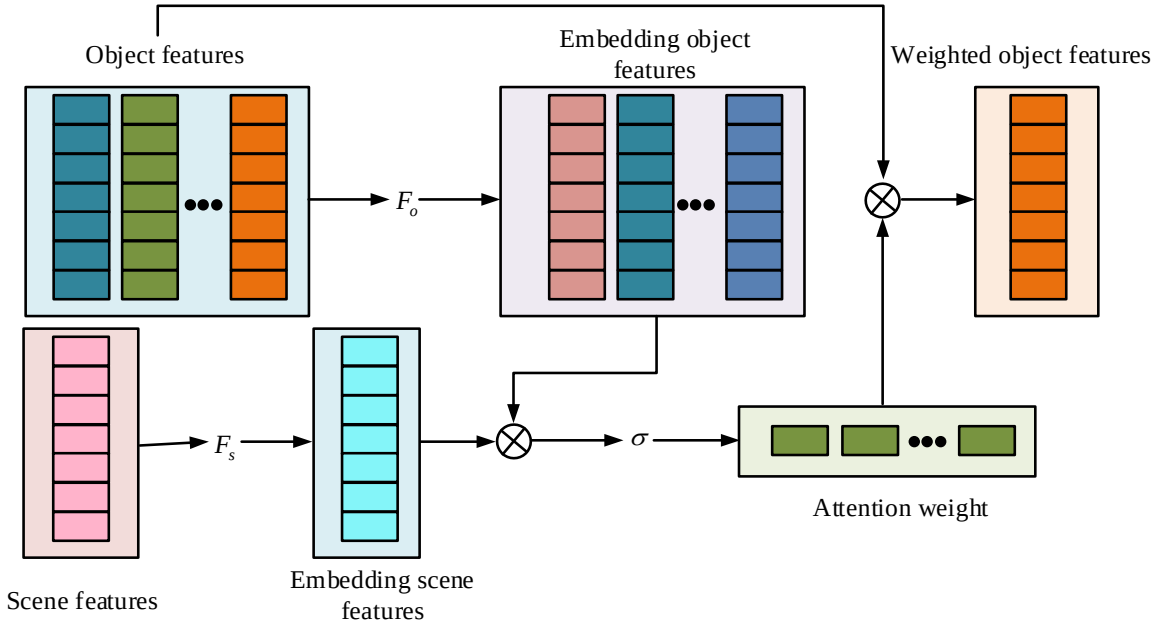

Fig. 5. Scenario-based attention mechanism.

In Eq. (16), the attention weight is a_i, and the activation function is σ. The embedded scene feature input is F_s, and the scene feature is f_{sce}. The embedded object feature is $o_i^{'}$, and the emotionally augmented object feature is F_o. The weighted object feature calculation is shown in Eq. (17).

$$f_{obj} = \sum_{i=1}^{M} a_i o_i^{'} \tag{17}$$

In Eq. (17), the weighted object feature is f_{obj} and the number of objects is M. The emotion feature after cascading the scene and the weighted object feature is shown in Eq. (18).

$$f_{emo} = Concate(f_{sce}, f_{obj}) \tag{18}$$

In Eq. (18), the cascaded sentiment feature is f_{emo}. The study interacts and fuses the scene and object features via AM to obtain weighted object features. Next, the scene and object features are cascaded to obtain the emotion features, which are used for subsequent emotion prediction along with a learnable weight matrix. The results of the sentiment prediction are compared with the sentiment class labels in the dataset and the network is optimized by CE loss.

III. Results

A. IEPN-MIF based Application Analysis

The experiments are performed on a computer based on the Pytorch framework with an Intel(R) Xeon (R) CPU E5-2640 2.40 GHz and an NVIDA GeForce GTX TITAN GPU (12G RAM) with Linux as the operating system. The experimental environment provides powerful computational capabilities to support the analysis and processing of large-scale datasets, which provides a good foundation for the training and testing of deep learning models. Three datasets are used for the experiments, namely, the FI dataset, the IAPS dataset and the ArtPhoto dataset. Table I displays the breakdown of the dataset. 2000 photos are chosen at random as the training set and 1000 images as the test set from the IAPS dataset. 2000 photographs are chosen at random as the test set and 6000 images as the training set from the Artphoto dataset. 15,000 photos are chosen at random as the training set and 7,000 images as the test set from the F1 dataset.

The experiments compare CNN-LSTM, LSTM, convolutional neural network (CNN), and the semantic emotion model proposed in 2023, and the various methods are applied in the small-scale dataset in order to validate the efficacy of the research proposed image emotion prediction network with multi-incentive fusion (IEPN-M). The accuracy and recall in IAPS are shown in Fig. 6.

Fig. 6 (a) shows the accuracy comparison in the dataset IAPS, the average accuracy of the study proposed IEPN-MIF is 80.3%, which improves about 18.6%, 29.5%, 25.6%, and 15.3% compared to the CNN-LSTM, LSTM, CNN models, and semantic sentiment models, respectively. Fig. 6 (b) shows the recall comparison in the dataset IAPS, and the average recall of the proposed IEPN-MIF is 83.2%, which is improved by about 10.3%, 14.3% 15.6%, and 9.8% compared to CNN-LSTM, LSTM, CNN model and semantic sentiment model, respectively. The results show that the multi-stimulus fusion methods have better performance in IE prediction tasks. The accuracy comparison of different methods on Artphoto and F1 datasets is shown in Fig. 7.

TABLE I. Dataset Division

Data set	Training set	Test set	Total number
IAPS	20000	1000	3000
Artphoto	6000	2000	8000
FI	15000	7000	23000

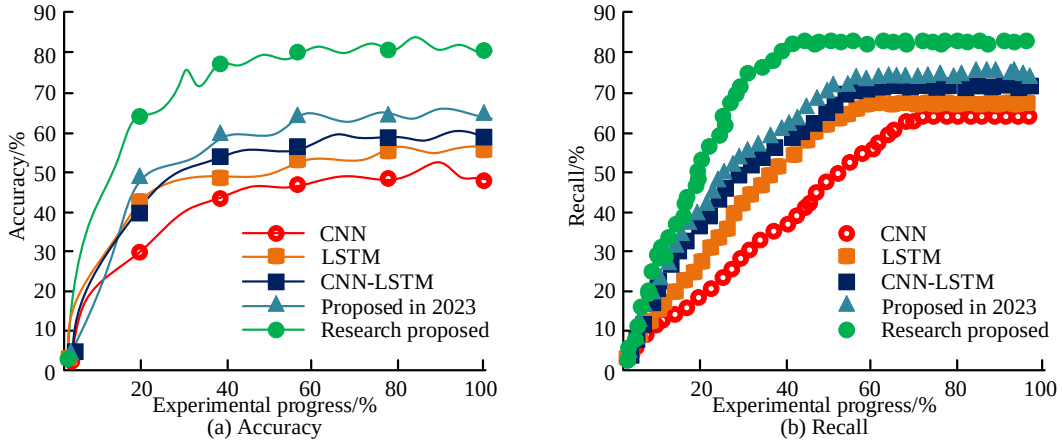

Fig. 6. Accuracy and recall in the dataset IAPS.

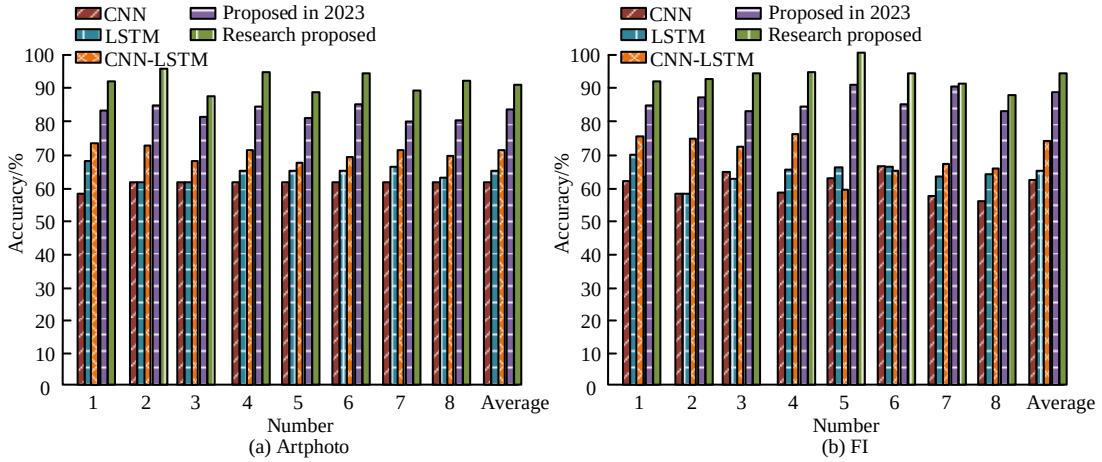

Fig. 7. Accuracy in the Artphoto and FI datasets.

TABLE II. RESULTS OF IMAGE EMOTION PREDICTION NETWORK ABLATION EXPERIMENTS FOR MULTI-EXCITATION FUSION

Global network		Semantic network		Expression network	Accuracy
RGB	Y	LSTM	FCL		
×	√	×	×	×	59.3%
√	×	×	×	×	67.2%
√	×	√	×	×	71.6%
√	×	×	√	×	62.7%
√	×	√	×	√	91.3%

In Fig. 7, 1-8 denote eight emotions, respectively. Fig. 7 (a) shows the accuracy comparison in the Artphoto dataset, and the average accuracy of IEPN-MIF is 91.3%, which improves about 19.6%, 23.8%, 27.8%, and 5.7% compared to CNN-LSTM, LSTM, CNN model and semantic sentiment model, respectively. Fig. 7(b) shows the accuracy comparison in the FI dataset, and the average accuracy of the IE prediction method with multi-incentive fusion is 93.6%, which improves about 20.7%, 30.1%, 27.5%, and 4.7% compared to CNN-LSTM, LSTM, CNN model, and semantic sentiment model, respectively. The outcomes demonstrate the increased accuracy and greater potential of the multi-excitation fusion IE prediction approach in the emotion identification domain. Table II displays the outcomes of the IEPN-MIF ablation trials.

In Table II, the global, semantic, and expression networks each contribute to sentiment prediction to varying degrees. When the three sub-networks act alone, the global network has the highest accuracy of 67.2%, indicating the strong representational power of global features. In the absence of color features, the three-branch network outperforms the two-branch network only to a lesser extent, while the three-branch network considering color features performs superiorly. The results show that by setting up the LSTM layer and the FCL for comparison, the LSTM layer structure can better mine the semantic information between different objects, and thus predict emotions more accurately. The addition of the expression network further improves the accuracy of the network, with a final accuracy of 91.3%. The results of the impact of hyperparameters on the network performance are shown in Fig. 8.

Fig. 8 (a) shows the effect of loss weights on network performance, and the accuracy of IEPN-MIF first increases and then decreases as the loss weights increase. The highest accuracy of 75.6% is achieved when the loss weight is 1.0. Fig. 8 (b) displays the effect of the object frames on the object performance, with the increase of the number of object frames, the accuracy of IEPN-MIF gradually increases and then decreases and stabilizes. When the object frame data is 10, the network has the highest accuracy rate of 76.5%. In conclusion, the model performs best when the IEPN-MIF loss weight and object frame count are set to 1.0 and 10, respectively.

B. Application Analysis of Image Emotion Classification based on Associative Reasoning

To validate the performance of the SOLVER network proposed in the study, the experiments use Faster R-CNN, GCN and the IE correlation analysis model based on hierarchical graph convolutional network proposed in 2023 as comparisons. The comparison of the accuracy of the different methods on the IAPS and Artphoto datasets is shown in Fig. 9.

Fig. 9 (a) shows the accuracy comparison in the IAPS dataset, where the average accuracy of SOLVER network is 92.1%, which is an improvement of about 29.8%, 25.3, and 7.4% compared to the Faster R-CNN, GCN, and IE correlation analysis models, respectively. Fig. 9 (b) shows the accuracy comparison in Artphoto dataset, the average accuracy of SOLVER network is 91.8%, which improves about 32.6%, 31.7, and 3.7% compared to Faster R-CNN, GCN and IE correlation analysis model, respectively. This indicates that the SOLVER network is able to recognize objects and scenes in images more accurately because of the more optimized inference method used in the SOLVER network. The results of the ablation experiments and hyperparametric analysis of the SOLVER network are shown in Fig. 10.

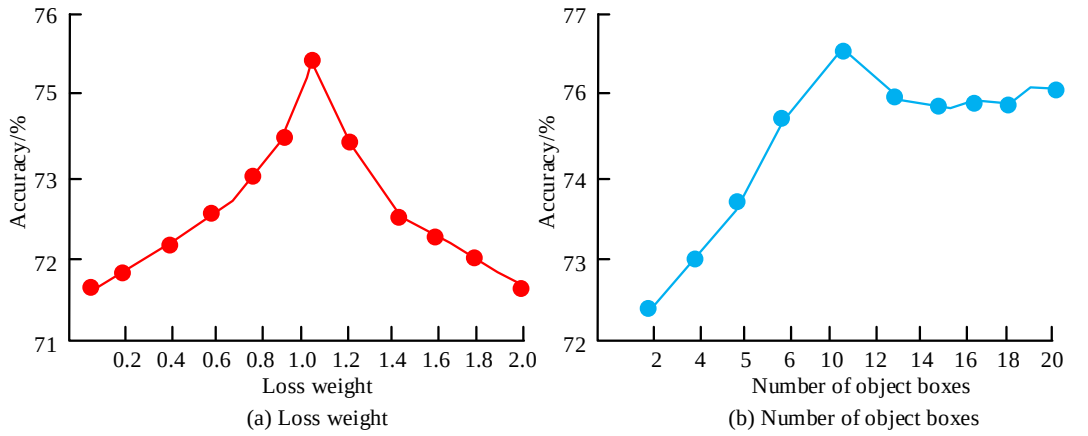

Fig. 8. Results of the influence of hyperparameters on network performance.

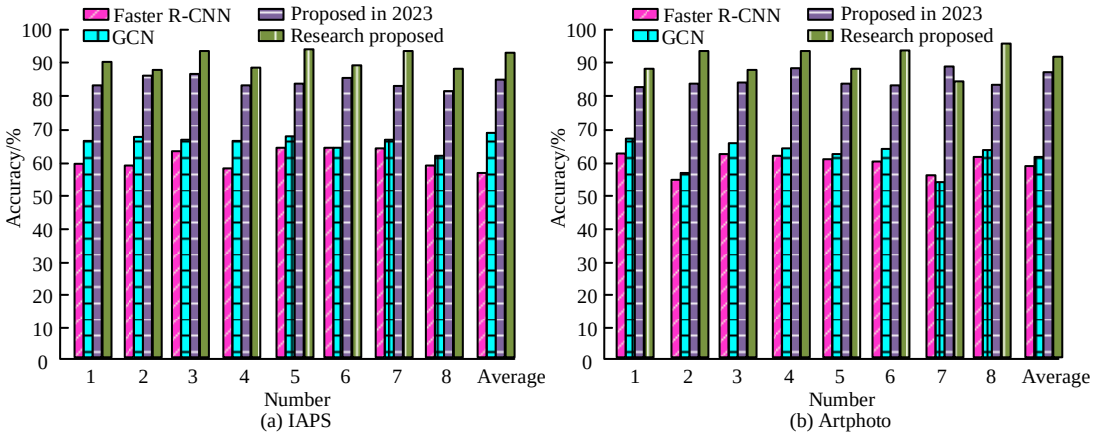

Fig. 9. Comparison of accuracy in the IAPS and Artphoto datasets.

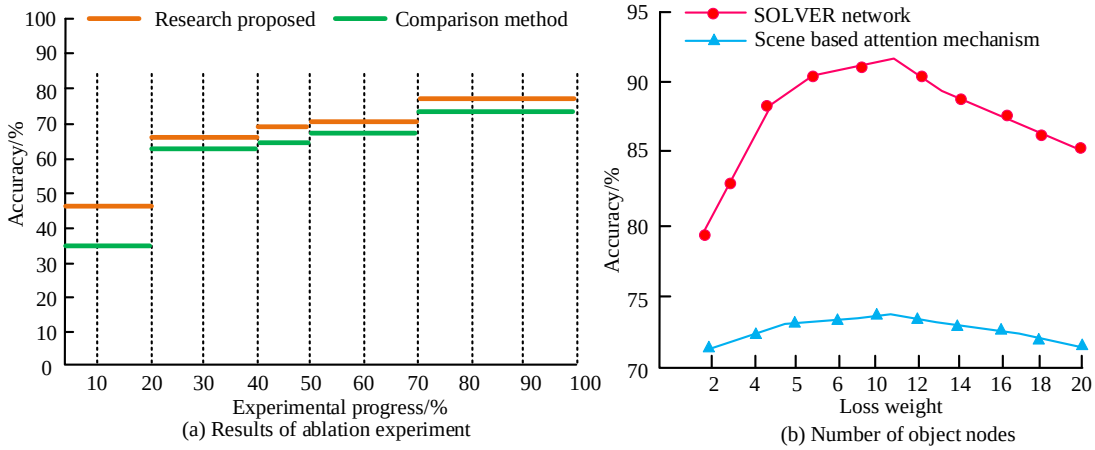

Fig. 10. Results of the ablation experiments and the hyperparameter analysis of the SOLVER network.

The results of the ablation tests are displayed in Fig. 10(a). In the 0%–10% phase, several objects result in a greater performance improvement than a single item. Two separate parameter embedding functions perform better than one embedding function in the 10%–40% stage. The mask operation can enhance performance even more in the 40%–50% range. Scene branching and scene-based AM both significantly contribute to emotion classification. In the 50%-70% stage, scene features direct the fusion process of object features. The results of the hyperparameter analysis are displayed in Fig. 10(b), where the accuracy of the scene-based AM and SOLVER network steadily rises as the number of object nodes increases until stabilizing. The scene-based AM and SOLVER networks get the maximum accuracy rates of 73.6% and 91.9%, respectively, when there are eleven object nodes. Therefore, the number of object nodes should be set to 11 to ensure network performance.

IV. Discussion

The proposed network image sentiment analysis method demonstrates high performance. This method fully utilized the advantages of ML in automatically recognizing image features, and deeply analyzed emotional expression through the idea of associative reasoning, achieving significant results in the field of sentiment analysis. In different emotional scenarios, this method could accurately identify various emotions such as joy, anger, sadness, and happiness, and achieve efficient emotional judgment in complex backgrounds. This was due to the fact that ML algorithms can effectively extract image features and improve the accuracy of emotion recognition. Meanwhile, the introduction of associative reasoning enabled this method to combine contextual information for more detailed analysis and judgment of emotional expression, thereby demonstrating high performance in different emotional scenarios. However, this fusion method also had certain limitations. Firstly, due to the ambiguity and diversity of emotional expression in network images, ML models might find it difficult to fully capture all key features during training, resulting in uncertainty in sentiment analysis results. Second, some environmental factors present in the associative reasoning process might affect the accuracy of sentiment analysis, resulting in results that are somewhat limited by human cognition. In addition, with the continuous changes in the expression of emotions in network images, ML and associative reasoning models needed to be constantly updated and optimized to adapt to new emotional trends.

These results are not limited to the dataset used, and the principles and methods behind them are applicable to other image and video datasets with rich emotional information. For example, in areas such as social media analysis, movie sentiment scoring, and advertising effectiveness evaluation, IEPN-MIF networks can accurately predict the emotional content of images by extracting emotional features from multiple sources, such as colors, objects, and faces. The SOLVER network can be applied to intelligent surveillance systems to support safety warnings and behavior understanding by analyzing the emotional relationship between scenes and objects. In the future, these networks are expected to be widely used in practical applications such as emotional interactive robots and personalized recommendation systems, thereby enhancing user experience and system intelligence.

V. Conclusion

To achieve accurate recognition and classification of IE, the study proposed an IEPN-MIF based network which was capable of utilizing different emotional incentives, including color, object, and face. It can achieve accurate prediction of emotions through deep learning techniques. Secondly, the study proposed SOLVER network to classify the emotion of images by fusing scene features and object features. The outcomes indicated that the average accuracy of the proposed IEPN-MIF in the dataset IAPS was 80.3%, which was improved by about 18.6%, 29.5%, 25.6%, and 15.3% compared to the CNN-LSTM, LSTM, CNN models and the semantic sentiment model, respectively. As the loss weight increased, the accuracy of IEPN-MIF first increased and then decreased. The highest accuracy of 75.6% was achieved when the loss weight was 1.0. The average accuracy of the SOLVER network in the IAPS dataset was 92.1%, which improved about 29.8%, 25.3, and 7.4% compared to the Faster R-CNN, GCN and IE correlation analysis models, respectively. Scene-based AM can effectively fuse scene and object features and classify emotions. Finally, the results of ablation experiments and hyperparameter analysis indicated that the number of object nodes should be set to 11 to ensure network performance. This method had great potential in practical applications, such as in the fields of intelligent interaction, psychology, and advertising marketing, where more accurate emotion recognition and classification could be achieved by analyzing IEs. Meanwhile, the scene based AM could effectively integrate scene and object features, and classify emotions, which provided the possibility for further improving the accuracy of IE recognition. The shortcoming of this research was that only datasets were used for the experiments. Therefore, the results only reflected the performance on these datasets. Future research directions can revolve around the following aspects. First, explore more advanced neural network architectures. Although CNN and LSTM networks have achieved good results in image sentiment analysis, the latest architectures such as Transformer AMs may provide better performance. Second, more effective loss functions should be developed, such as contrast loss or cosine similarity loss. Future research should test the methods on more diverse datasets to ensure their generalizability. Finally, consider multimodal sentiment analysis. Image sentiment analysis typically focuses only on visual content, and combining modal information, such as text associated with images or background audio, can improve analysis accuracy. With this dependency, future research can further expand the field of image sentiment analysis, develop more accurate and robust methods for understanding the emotional content of online images, and make greater contributions to the development of artificial intelligence technology.

References

[1] Hosseinalipour A, Ghanbarzadeh R. A novel metaheuristic optimisation approach for text sentiment analysis. International journal of machine learning and cybernetics, 2023, 14(3):889-909.

[2] Liu X, Zhang Z, Zhang G. Using improved feature extraction combined with RF-KNN classifier to predict coal and gas outburst. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2023, 44(1):237-250.

[3] Sun X, Cai Z. Research on an Eye Control Method Based on the Fusion of Facial Expression and Gaze Intention Recognition. Applied Sciences, 2024, 14(22): 10520-10542.

[4] Bhaumik G, Govil M C. SpAtNet: A spatial feature attention network for hand gesture recognition. Multimedia Tools and Applications, 2024, 83(14): 41805-41822.

[5] Sun Y, Guo Q, Zhao S, Chandran K, Fathima G. Context-aware augmented reality using human-computer interaction models. Journal of Control and Decision, 2024, 11(1): 1-14.

[6] Halim Z, Waqar M, Tahir M. A machine learning-based investigation utilizing the in-text features for the identification of dominant emotion in an email. Knowledge-Based Systems, 2020, 208(15):106443-106459.

[7] Britzolakis A, Kondylakis H, Papadakis N. A Review on Lexicon-Based and Machine Learning Political Sentiment Analysis Using Tweets. International Journal of Semantic Computing, 2021, 14(4):517-563.

[8] Alasmari W A, Abdelhafez H A. Twitter Sentiment Analysis for Reviewing Tourist Destinations in Saudi Arabia Using Apache Spark and Machine Learning Algorithms. Journal of computer sciences, 2022, 18(3):210-221.

[9] Chirgaiya S, Sukheja D, Shrivastava N, Rawat R. Analysis of sentiment based movie reviews using machine learning techniques. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021, 41(5):5449-5456.

[10] Kumar S, Gahalawat M, Roy P P, Dogra D P. Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning. Electronics, 2020, 9(2):374-387.

[11] Liewlom P. Alternative Rule Reasoning: Association Rule Tree Reasoning with a Constraining Rule Ascertained using a Reasoning Framework in 2D Interestingness Area. IAENG Internaitonal journal of computer science, 2021, 48(3):619-633.

[12] Li C, Yang X, Yin K, Chang Y, Wang Z, Yin G. Pedestrian re-identification based on attribute mining and reasoning. IET Image Processing, 2021, 15(11):2399-2411.

[13] Lin Y R, Lee C H, Lu M C. Robust tool wear monitoring system development by sensors and feature fusion. Asian Journal of Control: Affiliated with ACPA, the Asian Control Professors, Association, 2022, 24(3):1005-1021.

[14] Zhang L, Li Y, Yi J, Wang J. Research On The Application Of Network Intrusion Feature Extraction In Power Network. International Journal of Autonomous and Adaptive Communications Systems, 2021, 14(4):342-353.

[15] Guo B, Zhang Y, Gao J, Li C, Hu Y. SGLBP: Subgraph-based Local Binary Patterns for Feature Extraction on Point Clouds. Computer Graphics Forum: Journal of the European Association for Computer Graphics, 2022, 41(6):51-66.

[16] Mutanov G, Karyukin V, Mamykova Z. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. Computers, Materials and Continua, 2021, 69(1):913-930.

[17] Kelvin Leong, Anna Sung. An Exploratory Study of How Emotion Tone Presented in A Message Influences Artificial Intelligence (AI) Powered Recommendation System. Journal of Technology & Innovation, 2023, 3(2): 80-84.

[18] Zhao Y, Guo M, Sun X, Chen X, Zhao F. Attention-based sensor fusion for emotion recognition from human motion by combining convolutional neural network and weighted kernel support vector machine and using inertial measurement unit signals. IET signal processing, 2023, 17(4): 12201-12212.

[19] Ou Z, Wang H, Zhang B, Liang H, Hu B, Ren L, Liu Y, Zhang Y, Dai C, Wu H, Li W, Li X. Early identification of stroke through deep learning with multi-modal human speech and movement data. Neural Regeneration Research, 2024, 20(1): 234-241.

[20] Purohit J, Dave R. Leveraging Deep Learning Techniques to Obtain Efficacious Segmentation Results. Archives of Advanced Engineering Science, 2023, 1(1):11-26.

[21] Chinthamu N, Karukuri M. Data Science and Applications. Journal of Data Science and Intelligent Systems, 2023, 1(1): 83-91.

[22] Shanqing Z, Yujie C, Yiheng M, Jianfeng L, Li L, Rui B. A multi-level feature weight fusion model for salient object detection. Multimedia Systems, 2023, 29(3): 887-895.