IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 3, June, 2024, pp. 1163-1172
Journal homepage: http://iieta.org/journals/ts

Analysis and Emotion Recognition of Educational Network New Media Images Based on Deep Learning

Yuhan Zeng

College of Humanities and Arts, Hunan International Economics University, Changsha 410205, China

Corresponding Author Email: 2012230011@students.stamford.edu

https://doi.org/10.18280/ts.410306

Received: 7 January 2024
Revised: 22 April 2024
Accepted: 19 May 2024
Available online: 26 June 2024

Keywords:
deep learning, Educational Network New Media (ENNM) images, content annotation, emotion recognition, feature re-calibration, cyclic structural representation

ABSTRACT

In today's fast-paced information technology landscape, Educational Network New Media (ENNM) has become a crucial tool in the education sector, with educational content increasingly presented to students in the form of images and videos. Effectively analyzing and recognizing the information and emotional states in these images is key to improving educational quality and enhancing student learning experiences. Existing research demonstrates that deep learning techniques have achieved significant results in image content analysis and emotion recognition. However, traditional image content annotation methods often overlook the re-calibration of features when dealing with ENNM images, resulting in less accurate annotation outcomes. Additionally, current emotion recognition methods primarily focus on categorizing emotional types, neglecting the recognition of emotional intensity and subtle changes, which falls short of the high precision demands in educational settings. This paper proposes two innovative methods: first, an ENNM image content annotation method based on feature re-calibration, which enhances the accuracy and robustness of image content annotation by incorporating feature re-calibration techniques; second, an ENNM image emotion recognition method based on cyclic structural representation, which achieves fine-grained emotion recognition by constructing a progressive cyclic loss function that integrates emotion intensity and polarity. This research not only addresses the shortcomings of existing methods but also provides educators with more accurate and detailed tools for image content and emotion analysis, offering significant theoretical and practical value.

1. INTRODUCTION

In today's rapidly developing information technology era, educational network new media has become an important tool and resource in the field of education, with an increasing amount of educational content being presented to students in the form of images and videos [1-4]. With the widespread application of these visual contents, effectively analyzing and recognizing the information and emotional states in ENNM images has become a key issue for improving educational quality and student learning experience [5-8]. By using deep learning techniques to analyze and recognize the content and emotions of ENNM images, educators can better understand students' emotional reactions and provide strong support for personalized teaching.

Related research shows that image content analysis and emotion recognition based on deep learning have achieved significant results in multiple fields. In the field of education, accurate image content analysis can automatically annotate educational resources, providing teachers and students with efficient retrieval and usage methods [9-11]; while emotion recognition can help teachers understand students' emotional states in real-time, adjust teaching strategies timely, and improve teaching effectiveness [12-14]. Therefore, in-depth research on the content analysis and emotion recognition of ENNM images has important theoretical value and practical significance.

Although existing research methods have achieved preliminary results in image content analysis and emotion recognition, there are still some shortcomings. Firstly, traditional image content annotation methods often ignore the re-calibration of features when dealing with ENNM images, resulting in inaccurate annotation results [15-18]. Secondly, in terms of emotion recognition, most existing methods only focus on the judgment of emotion categories, ignoring the recognition of emotion intensity and subtle changes, which fails to meet the high-precision requirements of emotion analysis in educational scenarios [19-21]. Therefore, a more precise and refined analysis method is urgently needed to improve the effectiveness of content and emotion recognition of ENNM images.

The main research content of this paper is divided into two parts: one is the ENNM image content annotation method based on feature re-calibration, and the other is the ENNM image emotion recognition method based on cyclic structural representation. In terms of content annotation, by introducing feature re-calibration techniques, the accuracy and robustness of image content annotation are improved; in terms of emotion recognition, by constructing a progressive cyclic loss function and combining emotion intensity and polarity, fine-grained

emotion recognition is achieved. This research not only addresses the shortcomings of existing methods but also provides educators with more accurate and detailed tools for image content and emotion analysis, offering significant theoretical and practical value.

2. ENNM IMAGE CONTENT ANNOTATION METHOD BASED ON FEATURE RE-CALIBRATION

In terms of improving the performance of ENNM image content annotation, although there has been a large amount of research focused on spatial dimension optimization, this paper adopts an innovative approach by conducting in-depth research on the relationships between feature channels to achieve feature re-calibration. Unlike other image feature re-calibration, the feature re-calibration of ENNM image content focuses more on the semantic understanding of educational content and the extraction of relevant features, ensuring that the annotation can more accurately reflect the educational information in the image. To this end, this paper proposes an ENNM image content feature re-calibration method based on SENet, by assigning different weights to different channels, thereby enhancing important features related to educational content and suppressing irrelevant or interfering features, thus improving the accuracy and efficiency of image content annotation.

For the application scenario of ENNM image content annotation, the feature re-calibration method based on SENet can better serve the precise extraction of educational information. Its annotation process particularly focuses on the semantic relevance of educational content by enhancing important features related to educational information and suppressing irrelevant or interfering features, significantly differing from the general focus of image feature re-calibration. Specifically, first, it transforms the input feature map into a z×g×q feature map through a series of convolution and pooling operations; then, through the Squeeze operation $D1(.)$, compresses the feature map into a 1×1×z feature vector; then, through the Excitation operation $D2(, q)$, the dimension of the feature vector remains unchanged, but its value is recalculated. Finally, this feature vector with new values is weighted and fused with the original z×g×q feature map to obtain the feature re-calibration result. Figure 1 shows the SENet module schematic.

Specifically, the feature re-calibration method based on SENet can be divided into three parts: (1) Squeeze: First, perform global average pooling on the z2×g×q feature map to obtain a feature vector of size 1×1×z2, that is, compress each g×q feature map along the z2 direction to form a 1×1 real number sequence. Since the z2 g×q feature maps are a collection of local descriptors, these descriptors have global expressiveness for the entire image. In ENNM images, this process helps extract global information reflecting the overall educational scene, thereby forming a feature vector with a global receptive field. (2) Excitation: Next, use two fully connected layers to perform nonlinear transformations on the Squeeze results. As the network deepens, this operation makes the features more inclined towards a certain category, especially the categories related to educational content. In this way, the weights of the feature channels are optimized, making the features closely related to educational content more prominent, enhancing the network's ability to recognize and understand educational content. (3) Scale: Finally, based on the output of Excitation, obtain the weighted features. By multiplicatively weighting these important features with increased weights onto the original features, channel-level feature re-calibration is completed. This process ensures that in the content annotation of ENNM images, key educational information can be highlighted, and irrelevant or interfering information can be suppressed, thereby improving the accuracy and effectiveness of the annotation.

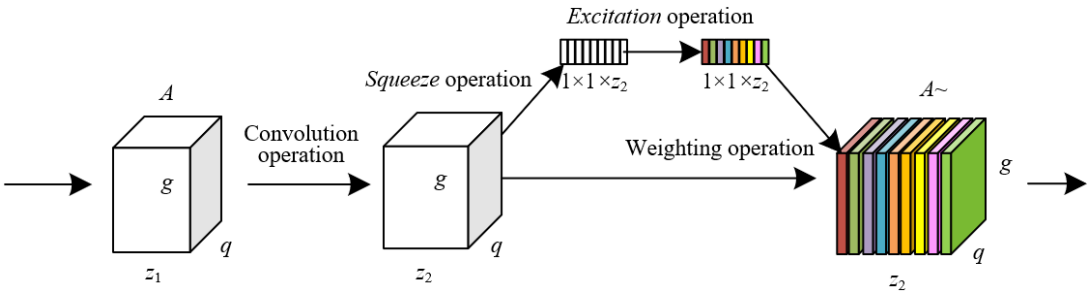

Figure 1. SENet module schematic

RefineDet is based on a feedforward convolutional neural network, generating a fixed number of bounding boxes and predicting the presence and scores of different category objects within these boxes. Then, the final classification results are produced through non-maximum suppression. RefineDet includes two inline modules and four feature fusion modules (TCB). The inline modules include the anchor refinement module (ARM) and the object detection module (ODM). ARM consists of a base network with the classification layer removed and auxiliary layers added, responsible for binary classification and regression, deleting refine anchors with negative confidence scores greater than 0.99. In the annotation of ENNM images, this step helps remove low-confidence irrelevant information and retain potential educational content areas. ODM uses the refine anchors from ARM as input and integrates the output of the TCB modules for further object detection, generating object scores and offsets. The TCB modules enhance the overall detection performance by discarding redundant information.

This paper chooses to add SE modules to the four-way features of the ARM and ODM modules in RefineDet, as shown in Figure 2. Since the method of adding SE modules to ARM and ODM is consistent, ARM is used as an example for explanation here, as shown in Figure 3. First, the SE module performs global average pooling on the z2×g×q feature map to obtain a 1×1×z2 feature vector, extracting globally expressive features. Then, two fully connected layers are used to perform nonlinear transformations on the feature vector, optimizing the weights of the feature channels, allowing the network to focus more on features related to educational content. Finally,

through multiplicative weighting, these optimized features are applied to the original features, completing channel-level feature re-calibration.

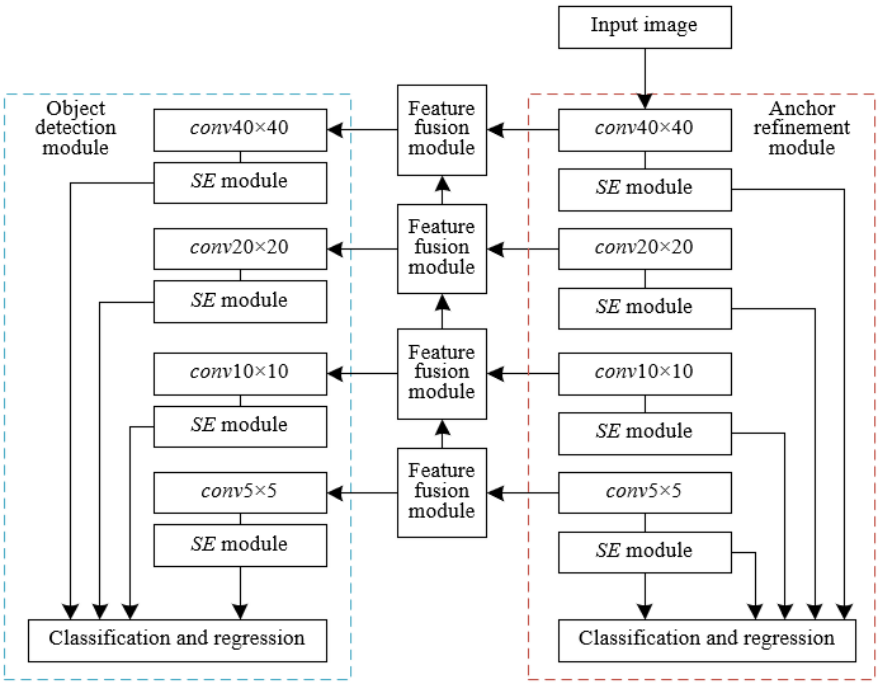

Figure 2. Schematic of added SE modules in ARM and ODM

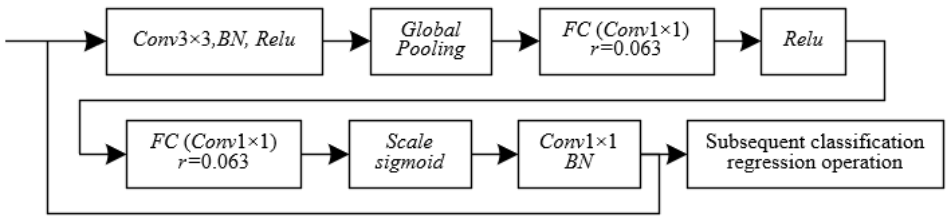

Figure 3. Schematic of added SE module in ARM

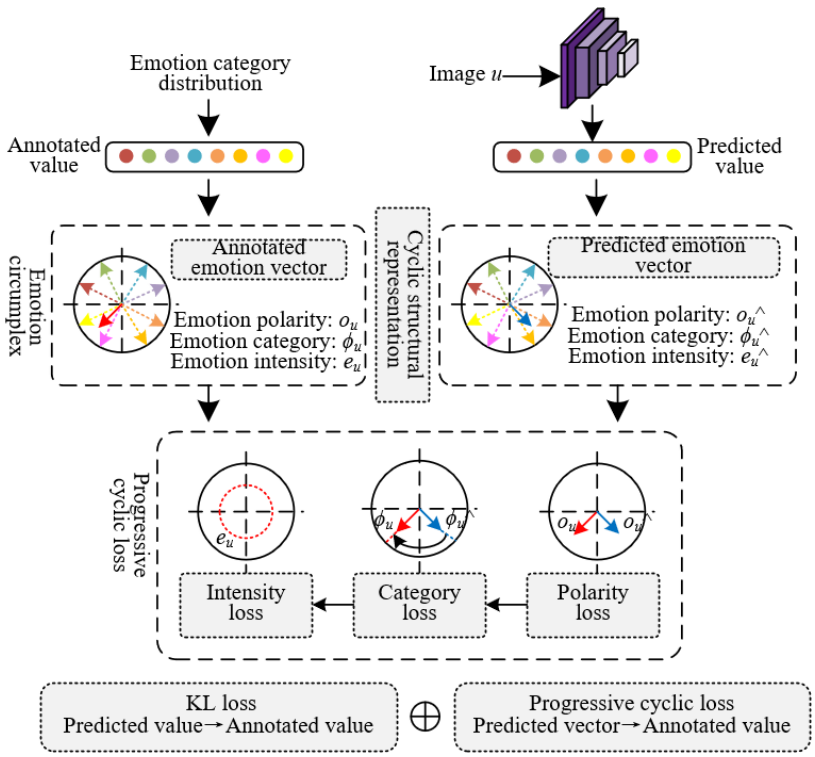

Figure 4. Schematic of the emotion recognition model for ENNM images

3. EMOTION RECOGNITION FROM ENNM IMAGES USING CYCLIC STRUCTURAL REPRESENTATION

Emotion recognition of ENNM images requires special attention to emotion features related to educational content. Through the RefineDet method based on feature re-calibration mentioned earlier, we have optimized the educational content features in the images, allowing the emotion recognition process to more accurately capture emotion expressions related to the educational context. Moreover, educational content often involves complex emotion expressions, such as motivation, encouragement, care, etc. Accurate recognition of these emotions is crucial for content annotation and user experience. To this end, this paper proposes a cyclic structure-based representation method for emotion recognition of ENNM images, analyzing and modeling image emotions by utilizing the inherent relationships between emotion categories. This method is proposed based on the circumplex model of affect in psychology, systematically guiding the learning of image emotion distribution by using the structure and characteristics of the emotion ring as prior knowledge. Furthermore, this paper proposes a progressive cyclic loss, combined with KL loss, to learn the differences between the predicted and labeled emotion distributions from coarse to fine. In this process, emotion characteristics are considered, enabling more refined class distribution learning. Figure 4 shows the schematic of the emotion recognition model for ENNM images.

3.1 Construction of the emotion circumplex

Inspired by psychological models, this paper constructs an emotion circumplex specifically for emotion recognition of ENNM images to learn image emotion distribution in a more specific and reasonable manner. The emotion circumplex provides a structured way to represent and learn these emotional states, allowing the emotion recognition process to better capture emotional expression features in educational contexts. By constructing the emotion circumplex, we can map any emotional state in ENNM images to an emotion vector r_u, which reflects not only the type of emotion but also its specific position and attributes on the circumplex. Assuming that emotion polarity, emotion category, and emotion intensity are represented by o_u, ϕ_u, and e_u respectively, we have:

$$r_u = (o_u, \varphi_u, e_u) \tag{1}$$

(1) Emotion Polarity o_u:

Unlike general emotion recognition tasks, emotion recognition in ENNM requires special attention to emotion polarity to ensure that educational content can appropriately convey emotional information, promoting learning and interaction. Drawing on Mikels' eight-category emotion dataset model, this paper subdivides emotions into two polarities: positive and negative. Specifically, in educational contexts, emotions such as encouragement, appreciation, satisfaction, and pleasure are classified as positive emotions, while frustration, sadness, disgust, and anger are classified as negative emotions. To align the structure of the emotion circumplex with the aforementioned emotion polarity division, this chapter evenly divides the emotion circumplex into two semicircles, one representing positive emotions and the other representing negative emotions. This division not only accurately reflects the emotion distribution in ENNM images but also, through the clear definition of emotion polarity, helps educators better understand and use emotional data to optimize teaching content and strategies.

$$o_u = \begin{cases} 0, \varphi_u \in \left[0, \frac{1}{2}\pi\right) \cup \left[\frac{3}{2}\pi, 2\pi\right) \\ 1, \varphi_u \in \left[\frac{1}{2}\pi, \frac{3}{2}\pi\right) \end{cases} \tag{2}$$

(2) Emotion Category ϕ_u:

To meet the complex emotional expression needs in ENNM, the emotion circumplex also includes compound emotions, which are distributed in the gaps between basic emotion vectors. These compound emotions can reflect more subtle emotional changes and interactions in educational scenarios, such as mixed emotional states that students may experience during the learning process. This detailed categorization of emotions can help educators gain a deeper understanding of students' emotional responses, thereby adjusting teaching strategies and improving educational outcomes. Therefore, the definition of emotion categories is crucial in the application of emotion recognition in ENNM images. To better maintain the circular structure on the emotion circumplex and meet the emotion recognition needs in educational contexts, this paper defines the emotion category $\phi_u \in [0,2\pi]$ through the polar angle in polar coordinates. This method can not only represent eight basic emotions: encouragement, appreciation, satisfaction, pleasure, frustration, sadness, disgust, and anger, but also cover more complex compound emotions. Assuming the number of emotion categories in the psychological model is represented by Z, the expression is:

$$\varphi^k = \frac{2k-1}{8}\pi, k \in \{1, 2, ..., Z\} \tag{3}$$

(3) Emotion Intensity e_u:

Emotional expressions in educational scenarios often have multi-layered complexity. For example, students may exhibit varying degrees of encouragement or frustration when facing learning tasks, and the intensity of these emotions provides important reference value for educators to adjust teaching methods and content. This paper further defines the emotion intensity of the eight basic emotion vectors to ensure sufficient subtlety in basic emotion recognition. Specifically, the radial coordinate in polar coordinates is used to represent the emotion intensity $e_u \in (0,1]$ of a specific emotion category ϕ_u. By setting the emotion intensity, we can more accurately capture and describe the emotional changes in ENNM images. For example, a photo showing a student successfully completing a task may have a very high emotion intensity, close to $e_u=1$, while a photo reflecting a student's slight confusion in learning may have a lower emotion intensity but still greater than 0. This detailed definition of emotion intensity helps educators promptly understand students' emotional states:

$$e^k = 1, k \in \{1, 2, ..., Z\} \tag{4}$$

(4) Similarity

In the application of emotion recognition in ENNM images, the definition of similarity between different emotions is also a key concept. This paper defines the similarity between

different emotions as the distance between their corresponding emotion vector polar angles. Specifically, image $u1$ and image $u2$ are considered to belong to the same emotion category if and only if $\phi_{u1}=\phi_{u2}$. This similarity definition is especially important in emotion recognition of ENNM images because emotional expressions in educational scenarios have specific contexts and meanings. For example, the confusion and slight anxiety exhibited by students in class may be very close emotionally, with a small polar angle difference, while there is a significant difference from pleasure and confidence, with a large polar angle difference. By clarifying the distance between emotion vector polar angles, we can more accurately identify and distinguish students' emotional expressions in different learning states.

(5) Additivity

According to psychological theory, emotions have additivity, where each compound emotion can be formed by the weighted combination of basic emotions. Therefore, it can be considered that emotional expressions in educational scenarios are often complex and multi-layered. For example, an image may simultaneously contain a student's curiosity and anxiety, and these two emotions can be combined into a compound emotional state through weighted addition, more accurately reflecting the student's real emotional experience in a specific educational context. This paper achieves the formation process from basic emotions to compound emotions through vector addition. In this way, the application of this additivity definition in emotion recognition of ENNM images not only helps identify complex emotional states but also provides educators with a systematic emotion analysis tool. For example, by weighting and combining different emotion vectors in an image, teachers can identify which images reflect students' positive emotions and which images reveal students' negative emotions. This detailed emotion analysis can help teachers promptly adjust teaching strategies and provide targeted emotional support.

3.2 Mapping of emotion vectors

Based on the three attributes and two features defined above, this paper proposes a systematic method for mapping the emotion distribution of any ENNM image into a compound emotion vector on the emotion circumplex. ENNM images usually contain emotional expressions of students in different learning scenarios, such as curiosity, anxiety, excitement, etc. Since the emotion vectors are defined in the polar coordinate system and the vector addition operation is defined in the Cartesian coordinate system, the proposed algorithm defines basic emotion vectors in the polar coordinate system and then merges the weighted basic emotion vectors in the Cartesian coordinate system, eventually obtaining a compound emotion vector with three attributes—emotion polarity, emotion category, and emotion intensity—represented in the polar coordinate system.

3.3 Progressive cyclic loss

The loss function for learning image emotion distribution is usually the Kullback-Leibler (KL) loss function. Suppose the annotated value of the emotion distribution of educational network new media images in the dataset is represented by f_u, and the predicted value of the emotion distribution is represented by $f_u^{\wedge}$. The features extracted from the emotion images using ResNet-50 are represented by d_u. The number of emotion images is represented by V, and the number of emotion categories in the dataset is represented by Z, then the function expressions are:

$$M_{JM} = -\frac{1}{V}\sum_{u=1}^{V}\sum_{k=1}^{Z} f_u(k) LN\hat{f}_u(k) \tag{5}$$

$$\hat{f}u_i(k \mid d_u, Q) = \frac{\exp(q_{u,k} d_u)}{\sum_{k=1}^{Z}\exp(q_{u,k} d_u)} \tag{6}$$

In this paper, the emotions of ENNM images are not a set of unrelated category labels but are distributed in a cyclic structure based on psychological models. To effectively utilize this prior knowledge, this paper further proposes a progressive cyclic loss, aiming to learn the emotion distribution from coarse to fine. The purpose of the progressive cyclic loss is to penalize the differences between the annotated emotion vector $r_u=(o_u,\phi_u,e_u)$ and the predicted emotion vector $r_u=(o_u^{\wedge},\phi_u^{\wedge},e_u^{\wedge})$. Specifically, this chapter progressively establishes constraints on the three attributes of the emotion vector: o_u, ϕ_u, and e_u. That is, the loss first ensures that the predicted emotion polarity of the ENNM image is consistent with its annotated emotion polarity through polarity loss, and then gradually refines the accuracy of the emotion category and emotion intensity of the ENNM image. This coarse-to-fine learning process allows the model to optimize step by step, thereby better adapting to the complex and subtle emotional features in ENNM images. Assuming the annotated value of emotion polarity is represented by o_u, and the predicted value of emotion polarity is represented by $o_u^{\wedge}$, the calculation formula is:

$$M_o = \frac{1}{V}\sum_{u=1}^{V}(o_u - \hat{o}_u)^2 \tag{7}$$

In the context of emotion recognition of ENNM images, students' emotional states not only affect their learning performance but also influence the classroom atmosphere and teachers' teaching strategies. Therefore, accurately identifying and distinguishing different emotion categories is particularly important. Hence, this paper further introduces category loss, which not only focuses on the differences between the predicted and annotated categories but also considers the distribution of these differences on the emotion circumplex. This is specifically achieved by measuring the difference in polar angles between the predicted emotion vector and the actual annotated emotion vector. This method leverages the emotion circumplex model, translating the similarity of emotion categories into geometric angle distances, thereby more intuitively reflecting the relationships between different emotion categories.

$$M_s = \frac{1}{V}\sum_{u=1}^{V}(\varphi_u - \hat{\varphi}_u)^2 \tag{8}$$

In the above formula, the closer the polar angles of the two emotion vectors, the more similar the emotional states they represent. However, merely considering emotion categories may not be sufficient. For example, suppose there are two ENNM images with the same emotion category but significantly different emotion intensities. In this case, it is

difficult to consider these two images as having the same emotional state. Therefore, in the construction of the progressive cyclic loss, we further introduce emotion intensity to more detailedly and accurately characterize emotional states. Specifically, this paper incorporates e_u as the confidence level of ϕ_u and o_u into polarity loss and category loss, namely:

$$M_{OZ} = \frac{1}{V}\sum_{u=1}^{V} e_u \left(\left(o_u - \hat{o}_u\right)^2 + \left(\varphi_u - \hat{\varphi}_u\right)^2 \right) \tag{9}$$

This paper combines the traditional KL loss with the proposed progressive cyclic loss, taking the weighted sum of the two as the loss function of the entire network. The KL loss measures the difference between the predicted distribution and the true distribution, while the PC loss refines the representation of emotional states by introducing emotion intensity and polar angles in the polar coordinate system. Assuming the hyperparameter balancing the two loss functions is represented by ω, the formula is:

$$M = (1-\omega) M_{JM} + \omega M_{OZ} \tag{10}$$

Through this method, the model can roughly identify emotion categories in the initial stage and perform finer adjustments and optimizations in subsequent stages by introducing emotion intensity and polarity. This progressive refinement process ensures the accuracy and sensitivity of emotion recognition, particularly in educational scenarios, enabling better capture of subtle changes in students' emotions.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Tables 1 and 2 present the comparison results of distribution metrics for content annotation of ENNM images on the training set. Our method shows significant advantages in all metric indicators. In Table 1, our method achieves the lowest value of 0.22 in Euclidean distance, significantly better than the Bayes Classifier (0.43) and SVM (0.54). In Manhattan distance and Minkowski distance, our method also achieved values of 0.78 and 0.67, respectively, outperforming other methods. In Bhattacharyya distance and Hellinger distance, our method stands out with values of 0.42 and 0.88, ranking first. Additionally, our method also excels in Intersection similarity and comprehensive ranking (Rank), with values of 0.72 and 1.2, respectively. In terms of accuracy (Acc.), our method achieves 0.71, higher than all other methods.

Table 1. Comparison-1 of distribution metrics for content annotation of ENNM images on the training set

Metrics	Problem Transformation Methods		Machine Learning Methods		LDL-Specific Methods		Our Method
	Bayes Classifier	*SVM*	*KNN*	*BP*	*LR*	*GP*	
Euclidean Distance	0.43(13)	0.54(14)	0.27(8)	0.35(11)	0.32(10)	0.36(12)	0.22(1)
Manhattan Distance	0.88(14)	0.88(13)	0.58(1)	0.81(8)	0.81(8)	0.85(12)	0.78(2)
Minkowski Distance	0.84(14)	0.82(13)	0.42(1)	0.74(10)	0.74(10)	0.81(12)	0.67(2)
Bhattacharyya Distance	1.87(13)	1.68(12)	3.24(14)	0.81(10)	0.65(7)	1.04(11)	0.42(1)
Hellinger Distance	0.62(13)	0.31(14)	0.78(7)	0.71(9)	0.77(8)	0.71(11)	0.88(1)
Intersection Similarity	0.48(13)	0.28(14)	0.63(5)	0.52(12)	0.61(9)	0.55(11)	0.72(1)
Rank	13.2(13)	13.2(13)	6(7)	11(11)	8.8(9)	11.6(12)	1.2(1)
Acc.	0.48(13)	0.38(14)	0.62(5)	0.57(11)	0.51(9)	0.51(12)	0.71(1)

Table 2. Comparison-2 of distribution metrics for content annotation of ENNM images on the training set

Metrics	CNN-Based Methods						Our Method
	DCNN	*LSTM*	*ACPNN*	*GAT*	*SSDL*	*E-GCN*	
Euclidean Distance	0.24(5)	0.24(5)	0.24(5)	0.23(4)	0.22(2)	0.22(2)	0.22(1)
Manhattan Distance	0.83(11)	0.77(5)	0.78(2)	0.78(2)	0.77(5)	0.77(5)	0.78(2)
Minkowski Distance	0.72(8)	0.71(7)	0.71(5)	0.71(3)	0.68(3)	0.68(3)	0.67(2)
Bhattacharyya Distance	0.71(8)	0.55(5)	0.52(6)	0.52(4)	0.45(3)	0.43(2)	0.42(1)
Hellinger Distance	0.73(9)	0.82(5)	0.81(6)	0.81(4)	0.84(3)	0.85(2)	0.88(1)
Intersection Similarity	0.61(7)	0.63(5)	0.66(7)	0.66(4)	0.67(3)	0.68(2)	0.72(1)
Rank	8(8)	5.2(6)	3.7(5)	3.7(4)	3.1(3)	2.8(2)	1.2(1)
Acc.	0.62(5)	0.62(5)	0.63(8)	0.63(4)	0.71(2)	0.68(3)	0.71(1)

Table 3. Comparison-1 of distribution metrics for content annotation of ENNM images on the test set

Metrics	Problem Transformation Methods		Machine Learning Methods		LDL-Specific Methods		Our Method
	Bayes Classifier	*SVM*	*KNN*	*BP*	*LR*	*GP*	
Euclidean Distance	0.53(13)	0.64(14)	0.27(7)	0.35(11)	0.29(7)	0.36(11)	0.22(1)
Manhattan Distance	0.87(7)	0.88(14)	0.58(1)	0.88(12)	0.85(11)	0.88(12)	0.83(3)
Minkowski Distance	0.74(6)	0.72(14)	0.42(1)	0.87(12)	0.76(11)	0.81(12)	0.77(2)
Bhattacharyya Distance	1.32(12)	1.38(13)	3.74(14)	1.21(10)	0.65(7)	1.18(10)	0.42(1)
Hellinger Distance	0.52(13)	0.51(14)	0.81(7)	0.71(11)	0.87(7)	0.71(11)	0.88(1)
Intersection Similarity	0.41(13)	0.48(14)	0.63(5)	0.57(9)	0.61(8)	0.56(11)	0.72(1)
Rank	10.2(12)	10.2(14)	5.8(6)	10.2(11)	8.4(9)	11.1(13)	1.4(1)
Acc.	0.45(13)	0.48(14)	0.72(7)	0.74(9)	0.71(10)	0.56(12)	0.77(1)

Table 4. Comparison-2 of distribution metrics for content annotation of educational network new media images on the test set

Metrics	CNN-Based Methods						Our Method
	DCNN	*LSTM*	*ACPNN*	*GAT*	*SSDL*	*E-GCN*	
Euclidean Distance	0.27(7)	0.24(5)	0.24(6)	0.23(3)	0.22(3)	0.23(2)	0.22(1)
Manhattan Distance	0.83(3)	0.87(3)	0.88(7)	0.88(2)	0.83(3)	0.87(7)	0.83(3)
Minkowski Distance	0.75(2)	0.71(6)	0.71(8)	0.71(2)	0.73(2)	0.78(8)	0.77(2)
Bhattacharyya Distance	0.66(7)	0.55(5)	0.52(6)	0.52(4)	0.52(3)	0.43(2)	0.42(1)
Hellinger Distance	0.81(7)	0.82(6)	0.81(5)	0.81(4)	0.84(3)	0.85(2)	0.88(1)
Intersection Similarity	0.57(10)	0.63(6)	0.66(7)	0.66(4)	0.67(3)	0.68(2)	0.72(1)
Rank	6(7)	5.2(5)	6.7(8)	3.1(3)	2.7(2)	3.8(4)	1.4(1)
Acc.	0.72(5)	0.72(7)	0.73(5)	0.73(3)	0.76(2)	0.78(3)	0.77(1)

Table 5. Ablation study of the proposed loss function on the training set

Metrics	*KL* Loss	*KL* Loss+ Polarity Loss	*KL* Loss+ Category Loss	*KL* Loss +Polarity Loss+ Category Loss	Our Loss Function
Euclidean Distance	0.225	0.221	0.221	0.221	0.221
Manhattan Distance	0.784	0.784	0.784	0.784	0.784
Minkowski Distance	0.689	0.678	0.689	0.689	0.685
Bhattacharyya Distance	0.425	0.435	0.425	0.425	0.402
Hellinger Distance	0.835	0.856	0.865	0.874	0.887
Intersection Similarity	0.665	0.687	0.715	0.701	0.702
Acc.	0.658	0.689	0.712	0.725	0.723

Table 6. Ablation study of the proposed loss function on the test set

Metrics	*KL* Loss	*KL* Loss+ Polarity Loss	*KL* Loss+ Category Loss	*KL* Loss +Polarity Loss+ Category Loss	Our Loss Function
Euclidean Distance	0.256	0.231	0.231	0.221	0.223
Manhattan Distance	0.845	0.854	0.835	0.832	0.835
Minkowski Distance	0.798	0.778	0.778	0.789	0.765
Bhattacharyya Distance	0.456	0.465	0.456	0.456	0.435
Hellinger Distance	0.832	0.897	0.875	0.875	0.879
Intersection Similarity	0.678	0.702	0.702	0.721	0.721
Acc.	0.735	0.756	0.775	0.798	0.789

In Table 2, our method also achieved the lowest values of 0.22, 0.78, and 0.67 in Euclidean distance, Manhattan distance, and Minkowski distance, respectively, showing high precision and robustness. In Bhattacharyya distance and Intersection similarity, our method achieved the best values of 0.42 and 0.72, especially significantly better in Bhattacharyya distance. Moreover, in comprehensive ranking (Rank) and accuracy (Acc.), our method ranked first with values of 1.2 and 0.71, respectively. In contrast, CNN-based DCNN methods and LSTM methods performed relatively poorly in most metrics, failing to achieve the best values in any indicator, demonstrating the limitations of traditional methods in new media image content annotation.

From the results of Tables 3 and 4, it can be seen that our proposed method performed excellently on multiple metric indicators, far surpassing other traditional methods. In Table 3, our method achieved the best values of 0.22, 0.83, and 0.77 in Euclidean distance, Manhattan distance, and Minkowski distance, respectively, showing high precision and robustness. In Bhattacharyya distance and Intersection similarity, our method achieved the best values of 0.42 and 0.72, significantly better than other methods. Moreover, in comprehensive ranking (Rank) and accuracy (Acc.), our method ranked first with values of 1.4 and 0.77, respectively. In contrast, traditional Bayes Classifier, SVM, and machine learning methods performed relatively poorly in most metrics, failing to achieve the best values in any indicator, demonstrating the limitations of traditional methods in new media image content annotation.

In Table 4, our method achieved the best values of 0.22, 0.83, and 0.77 in Euclidean distance, Manhattan distance, and Minkowski distance, respectively, showing high precision and robustness. Especially in Bhattacharyya distance and Intersection similarity, our method achieved the best values of 0.42 and 0.72, significantly better than other methods. Moreover, in comprehensive ranking (Rank) and accuracy (Acc.), our method ranked first with values of 1.4 and 0.77, respectively. In contrast, CNN-based DCNN methods and LSTM methods performed relatively poorly in most metrics, failing to achieve the best values in any indicator, demonstrating the limitations of traditional methods in new media image content annotation.

The results in Table 5 show that our proposed loss function outperformed other combinations in the ablation study on the training set. Specifically, our loss function showed similar performance in Euclidean distance, Manhattan distance, and Minkowski distance compared to other combinations, maintaining low levels and demonstrating high accuracy and stability. However, in Bhattacharyya distance and Hellinger distance, our loss function achieved the best values of 0.402 and 0.887, significantly better than other combinations, especially in Bhattacharyya distance. Additionally, in Intersection similarity and accuracy (Acc.), our loss function also achieved high values of 0.702 and 0.723, demonstrating strong overall performance. From the results in Table 6, it can be seen that our proposed loss function generally outperformed other combinations in the ablation study on the test set. Specifically, our loss function performed best in Euclidean distance and Minkowski distance, with values of

0.223 and 0.765, respectively, showing high accuracy and robustness. In Manhattan distance, our loss function achieved a result of 0.835, slightly inferior to the 0.832 of KL Loss + Polarity Loss + Category Loss, but still performing well. In Bhattacharyya distance, our loss function achieved the lowest value of 0.435, significantly better than other combinations. In Hellinger distance and Intersection similarity, our loss function achieved good results of 0.879 and 0.721, respectively. Finally, in the accuracy (Acc.) metric, our loss function achieved a high value of 0.789, second only to the 0.798 of KL Loss + Polarity Loss + Category Loss.

From the comparison results of annotated values and predicted values of the four dataset samples shown in Figure 5, it can be seen that the method proposed in this paper shows high accuracy and robustness in emotion distribution prediction. Specifically, in the emotion label dataset, the annotated and predicted values are close in the main emotion categories (2 and 3), especially the predicted value of category 2 being 0.58, close to the annotated value of 0.64. In the educational content dataset, the predicted value of category 3 is 0.66, although slightly lower than the annotated value of 0.8, the overall trend matches. In the multimodal dataset, the predicted value of category 3 is 0.42, close to the annotated value of 0.39, and the predicted value of category 4 is 0.28, also not far from the annotated value of 0.24. Finally, in the educational network new media dataset, the predicted value of category 3 is 0.6, close to the annotated value of 0.62, showing overall good performance.

Through the comparative analysis of the four datasets, it can be seen that the method proposed in this paper is effective and advantageous in emotion recognition and content annotation. Through feature re-calibration techniques and progressive cyclic loss functions, our method can accurately reflect the distribution of annotated values in the main emotion categories across different datasets. This indicates that our method not only performed well on a single dataset but also showed strong adaptability and stability across diverse datasets, further verifying its practicality and reliability in content annotation and emotion recognition of ENNM images. Overall, the research results of this paper provided new technical means and methods for this field, with important theoretical value and application prospects.

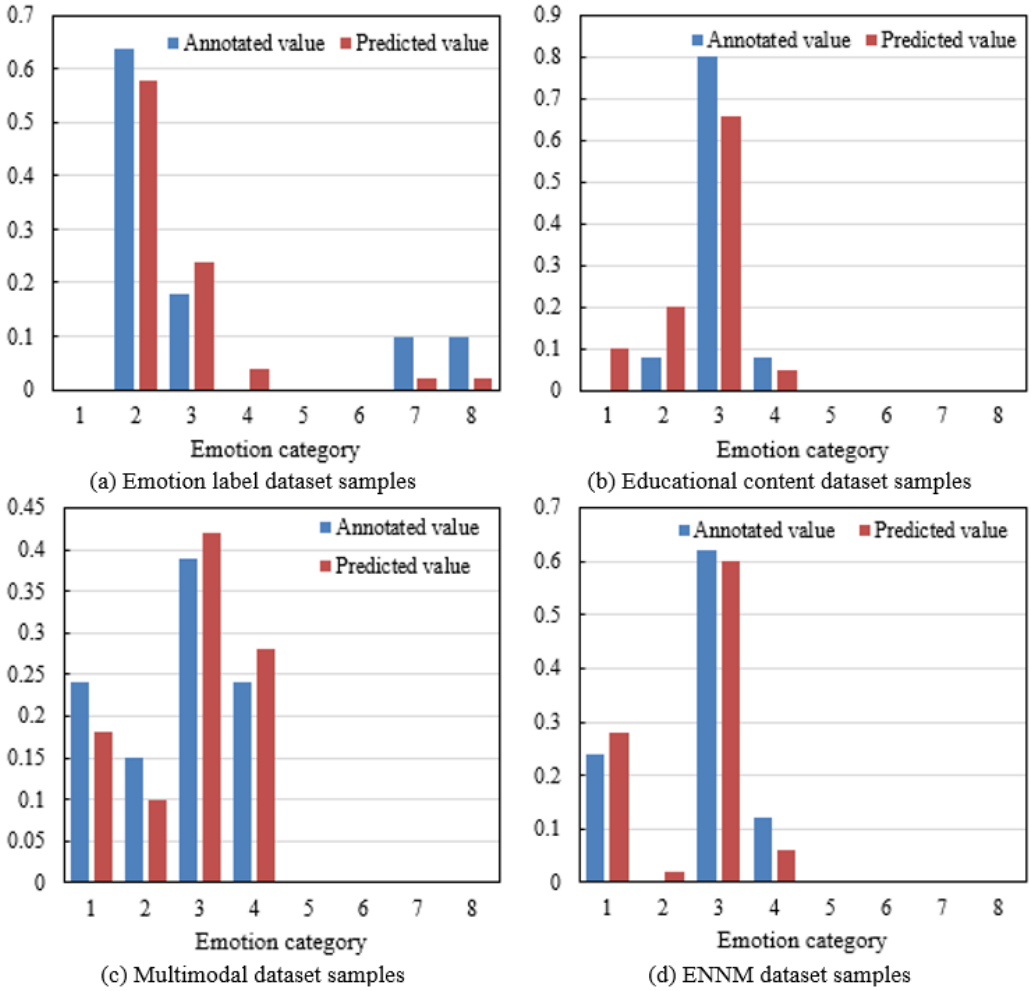

Figure 5. Visualization analysis of content annotation and emotion distribution of ENNM images

5. CONCLUSION

This paper's main research content is divided into two parts: one is the method of content annotation for ENNM images based on feature re-calibration, and the other is the emotion recognition method for ENNM images based on cyclic structural representation. In terms of content annotation, the introduction of feature re-calibration technology significantly improved the accuracy and robustness of image content annotation; in terms of emotion recognition, the construction of a progressive cyclic loss function, combined with emotion intensity and polarity, achieved fine-grained recognition of

emotional states. Experimental results show that the proposed method performed excellently on distribution metrics for both the training set and test set, particularly in key metrics such as Bhattacharyya distance, Minkowski distance, Intersection similarity, and accuracy, where it demonstrated significant advantages. Additionally, the results of ablation experiment further validated the effectiveness of feature re-calibration technology and cyclic structural representation in improving model performance. By visualizing the emotion distribution and content annotation results on various datasets, the accuracy and robustness of the proposed method in emotion recognition and content annotation are fully demonstrated.

This research provided new technical means and methods for content annotation and emotion recognition of ENNM images, with important theoretical value and practical application prospects. Feature re-calibration technology effectively enhanced the model's ability to capture image details and complex features, while cyclic structural representation, combined with emotion intensity and polarity, achieved more precise recognition of emotional states. However, the method still has some limitations, such as the need to improve prediction accuracy for certain specific emotion categories and the need to further optimize model performance when handling multimodal data. Future research directions can focus on the following aspects: firstly, further improving feature re-calibration technology and cyclic structural representation methods to enhance the recognition accuracy of complex emotional states; secondly, exploring methods for multimodal data fusion to improve the model's adaptability and generalization ability across different datasets; and thirdly, applying the proposed method to more practical scenarios to verify its effectiveness in different fields, and continuously optimizing and improving the model to meet practical needs. Through continuous deepening of research and technological innovation, the proposed method is expected to achieve greater breakthroughs in the field of content annotation and emotion recognition of ENNM images.

REFERENCES

[1] Zhang, C., Shen, X. (2024). Opportunities and challenges of new media technology for civic education of college students under the new normal. Applied Mathematics and Nonlinear Sciences, 9(1): 1-12. https://doi.org/10.2478/amns.2023.1.00250

[2] Sun, H. (2017). Research on the network development and the combination of theory and practice of political education in universities based on new media. Boletin Tecnico/Technical Bulletin, 55(15): 369-374.

[3] Zhu, S. (2017). Study on the current situation and effective countermeasures of college network ideological and political education based on the visual angle of new media. Boletin Tecnico/Technical Bulletin, 55(18): 393-399.

[4] Song, R., Liu, S., Zanuttini, J.Z. (2024). The effect of dance education on college students' artistic quality under the new media. International Journal of Web-Based Learning and Teaching Technologies, 19(1): 337969. https://doi.org/10.4018/IJWLTT.337969

[5] Liang, Y., Pan, F. (2024). Interactive experience design of traditional dance in new media era based on action detection. Computer-Aided Design and Applications, 21(S7): 241-255. https://doi.org/10.14733/cadaps.2024.S7.241-255

[6] Cui, L., Zhang, Z., Wang, J., Meng, Z. (2022). Film effect optimization by deep learning and virtual reality technology in new media environment. Computational Intelligence and Neuroscience, 2022(1): 8918073. https://doi.org/10.1155/2022/8918073

[7] Su, Y., Sun, W. (2023). Classification and interaction of new media instant music video based on deep learning under the background of artificial intelligence. The Journal of Supercomputing, 79(1): 214-242. https://doi.org/10.1007/s11227-022-04672-4

[8] Liu, H., Ko, Y.C. (2021). Cross-media intelligent perception and retrieval analysis application technology based on deep learning education. International Journal of Pattern Recognition and Artificial Intelligence, 35(15): 2152023. https://doi.org/10.1142/S0218001421520236

[9] Zhang, L.J., Wu, J.Z., Wei, J.X., Yu, X.Y., Yu, J., Yuan, B. (2023). Enhanced laboratory safety education through interactive applications of machine learning-boosted image processing technologies. Traitement du Signal, 40(6): 2623-2633. https://doi.org/10.18280/ts.400624

[10] Wang, Y., Haq, N.F., Cai, J., Kalia, S., Lui, H., Wang, Z.J., Lee, T.K. (2022). Multi-channel content based image retrieval method for skin diseases using similarity network fusion and deep community analysis. Biomedical Signal Processing and Control, 78: 103893. https://doi.org/10.1016/j.bspc.2022.103893

[11] Kumar, A., Saudagar, A.K.J., AlKhathami, M., Alsamani, B., Hasanat, M.H.A., Khan, M.B., Kumar, A., Singh, K.U. (2022). AIAVRT: 5.0 transformation in medical education with next generation AI- 3D animation and VR integrated computer graphics imagery. Traitement du Signal, 39(5): 1823-1832. https://doi.org/10.18280/ts.390542

[12] Haq, H.B.U., Akram, W., Irshad, M.N., Kosar, A., Abid, M. (2024). Enhanced Real-Time Facial Expression Recognition Using Deep Learning. Acadlore Transactions on AI and Machine Learning, 3(1): 24-35. https://doi.org/10.56578/ataiml030103

[13] Liang, S., Wu, D., Zhang, C. (2024). Enhancing image sentiment analysis: A user-centered approach through user emotions and visual features. Information Processing & Management, 61(4): 103749. https://doi.org/10.1016/j.ipm.2024.103749

[14] Zhu, Z., Zheng, X.Q., Ke, T.P., Chai, G.F. (2023). Emotion recognition in learning scenes supported by smart classroom and its application. Traitement du Signal, 40(2): 751-758. https://doi.org/10.18280/ts.400235

[15] Delacroix, T., Lomello, F., Schuster, F., Maskrot, H., Jacquier, V., Lapouge, P., Coste, F., Garandet, J.P. (2023). Measurement of powder bed oxygen content by image analysis in laser powder bed fusion. Materials & Design, 226: 111667. https://doi.org/10.1016/j.matdes.2023.111667

[16] Tang, D.Y.Y., Chew, K.W., Ting, H.Y., et al., (2023). Application of regression and artificial neural network analysis of Red-Green-Blue image components in prediction of chlorophyll content in microalgae. Bioresource Technology, 370: 128503. https://doi.org/10.1016/j.biortech.2022.128503

[17] Singh, S., Gandhi, M., Kar, A.K., Tikkiwal, V.A. (2023). How should B2B firms create image content for high social media engagement? A multimodal analysis.

Industrial Management & Data Systems, 123(7): 1961-1981. https://doi.org/10.1108/IMDS-08-2022-0470

[18] Tohgasaki, T., Aihara, S., Ikeda, M., et al. (2024). Investigation of stratum corneum cell morphology and content using novel machine‐learning image analysis. Skin Research and Technology, 30(2): e13565. https://doi.org/10.1111/srt.13565

[19] Zhu, T., Li, L., Yang, J., Zhao, S., Liu, H., Qian, J. (2023). Multimodal sentiment analysis with image-text interaction network. IEEE Transactions on Multimedia, 25: 3375-3385. https://doi.org/10.1109/TMM.2022.3160060

[20] Baoyue, H., Ashraf, H., Jhanjhi, N., Verma, S. (2024). Sentiment analysis system for image and text based social media data. In 2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC), Tandojam, Pakistan, pp. 1-7. https://doi.org/10.1109/KHI-HTC60760.2024.10482213

[21] Zhang, T., Zhou, G., Lu, J., Li, Z., Wu, H., Liu, S. (2024). Text-image semantic relevance identification for aspect-based multimodal sentiment analysis. PeerJ Computer Science, 10: e1904. https://doi.org/10.7717/peerj-cs.1904