Research Postprint on Image Emotion Feature Extraction Based on Deep Learning CNN Models

Authors: Li Zhiyi, Xu Hongkai, Duan Bin

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Image retrieval based on user emotions has become a research hotspot in machine learning. However, corpus data for image emotion feature annotation mostly originates from extraction of low-level image features, resulting in a monotonous and procedural image retrieval process. This paper proposes a deep learning-based algorithm for image emotion feature extraction that fuses low-level image features into high-level emotional semantics, providing a reference for implementing emotional semantic retrieval of images. [Method/Process] Using an improved convolutional network model, with color and texture of dataset images as input, emotional information is automatically extracted from images through multi-layer operations, and the emotion retrieval accuracy of the improved model is calculated via the backpropagation algorithm, thereby constructing an image emotion feature extraction model with high accuracy and low overfitting. [Results/Conclusion] By applying the improved convolutional neural network model, extraction of image emotion features is achieved, which improves retrieval accuracy by 10% compared to the original model.

Full Text

Preamble

Vol. 63 No. 11 June 2019 ChinaXiv Cooperative Journal

Research on Image Emotion Feature Extraction Based on Deep Learning CNN Models

Li Zhiyi[1], Xu Hongkai[1], Duan Bin[2]
[1]School of Economics and Management, South China Normal University, Guangzhou 510006
[2]School of Information and Optoelectronic Science and Technology, South China Normal University, Guangzhou 510006

Abstract

[**Purpose/Significance**] Image retrieval based on user emotions has become a hotspot in machine learning research. However, corpus data for image emotion feature annotation is mostly derived from low-level feature extraction, leading to a monolithic and stylized retrieval process. This paper proposes a deep learning-based algorithm for image emotion feature extraction that integrates low-level image features with high-level emotional semantics, providing a reference for emotional semantic retrieval of images.

[**Method/Process**] Using an improved convolutional network model, color and texture features of dataset images are taken as input. Emotional information is automatically extracted through multi-layer operations, and the emotion retrieval accuracy of the improved model is calculated via backpropagation to construct an image emotion feature extraction model with high accuracy and low overfitting.

[**Result/Conclusion**] The improved convolutional neural network model achieves image emotion feature extraction, improving retrieval accuracy by 10% compared to the original model.

Keywords: deep learning; image; emotional features; extraction; convolutional neural network
Classification Number: TP391.1
DOI: 10.13266/j.issn.0252-3116.2019.11.011

With the rapid development of social media platforms such as Weibo, forums, Zhihu, and Douban, an increasing number of people express opinions, share knowledge, and create content online. This user-generated content often carries subjective emotional tendencies, and sentiment analysis based on textual features and natural language processing has played a significant role in effectively capturing user sentiment. Today, with the widespread use of mobile social platforms, uploading photos to express emotions has become a popular form of communication. Analyzing, organizing, and mining emotional information from these images is crucial for hotspot detection and public opinion analysis [1]. However, the richness of visual information and the diversity of human subjective cognition make emotion-based image feature extraction particularly challenging.

From an emotional perspective, retrieving images based on image emotional semantics helps capture users' emotional states when viewing images, optimizes user emotion feature libraries, and improves emotion-based image retrieval systems. Currently, most image emotion feature annotation corpora are derived from low-level feature extraction, making the retrieval process monolithic and stylized. This paper attempts to propose an image emotion feature extraction algorithm based on the VGGNet-16 model that integrates low-level features with high-level emotional semantics to address existing limitations in emotional semantic feature extraction.

1. Literature Review

Research on user emotional features has matured in text-based sentiment analysis, where emotional feature extraction from user-generated text has become an important method for exploring user sentiment clues. For example, Ma Songyue et al. [2] used the ROSTCM6.0 sentiment analysis tool to extract sentiments from Douban movie reviews, segmenting user comments and calculating sentiment values for visualization to determine user tendencies and attach emotional labels to movies. Jiang Zhiyi et al. [3] used a self-built sentiment dictionary to segment Weibo content, calculating sentiment values to classify user tendencies as positive, negative, or neutral, thereby judging emotional evolution trends. The limitation of text-based sentiment analysis is its inability to handle non-textual data. With the rapid development of self-media, images have become a primary type of user-generated content and can better reflect user emotions than text [4], bringing image emotion analysis into research focus.

Image emotional semantic extraction has undergone two developmental stages: the first based on visual features, and the second represented by machine learning, especially deep learning, whose theory has been widely applied in image and video analysis, computer vision, speech recognition, and multimedia with great success [5]. In this stage, machine learning methods are applied to image emotional semantic feature extraction. Typical algorithm models include: (1) Deep Belief Network (DBN), which has achieved significant results in image classification by training each network layer step-by-step, mapping image features to different feature spaces, and establishing joint distributions between observed data and feature labels through probability distributions to improve classification accuracy [6]. (2) Deep Residual Network (DRN), which has attracted widespread attention in computer vision and cross-modal data processing. To address the "degradation problem" where accuracy quickly saturates and then declines with increased network depth, it uses identity mapping in added network layers to control training errors [7].

Comparatively, international research on image extraction started earlier and has made progress. J.A. Black et al. [8] conducted comparative experiments between topic indexing and emotion indexing, finding significant consistency in clustering results, reflecting the feasibility of image indexing and retrieval based on user emotions. K. Yoshida et al. [9] extracted color features to map user visual perception with image features for emotional semantic extraction. S.B. Cho et al. [10] described images using happiness, depression, and coolness, establishing an image emotion semantic extraction system. C. Colombo et al. [11] used pleasant, tense, relaxed, and dynamic to describe image emotional semantics, establishing mappings between images and emotions. S. Siersdorfer et al. [12] used color histograms to extract color features and SentiWordNet to score text information for sentiment extraction, finally using SVM classifiers to combine color features with sentiment, optimizing the mapping and improving matching efficiency.

Domestic scholars have also conducted related research. Huang et al. [16] noted that image emotional features originate from images, users, and environment, proposing a user-centered approach to image description and retrieval. Wang Shangfei et al. [17] extracted RGB distribution features and used radial basis function neural networks for image emotion feature extraction, forming a preliminary content-based emotion feature model. Wu Pin et al. [18] used support vector machines for small-sample automatic emotion annotation, designing an emotion-based image retrieval system. Wang Huaqiu et al. [19] used shape and color moments as fuzzy neural network inputs in image emotion semantic models, optimizing the mapping with genetic algorithms to improve efficiency.

Current research has transitioned from low-level feature-based extraction to deep learning-based methods, achieving significant progress. However, there remains substantial room for improvement in feature extraction models to enhance algorithm superiority. The emergence of facial emotion recognition technology presents new challenges in computer vision [20]. The particularity of human emotional information and the complexity of processing human emotions indicate that combining image emotion with human emotion in feature extraction methods still has vast development potential.

2. General Methods for Low-Level Image Feature Extraction

To extract image emotional features, low-level features must first be extracted to connect high-level emotional semantics and establish emotional characteristics.

2.1 Low-Level Semantic Features and Emotion

The most basic visual elements in images are color and texture. When observing images, people first receive these visual features, which then influence their emotional responses. Among visual features, color is the most significant and emotionally expressive feature, corresponding to different subjective user feelings [21]. Psychological experiments show that red corresponds to excitement, passion, progress, rage, and intensity; green corresponds to freshness, tranquility, nature, and fatigue. Since emotions contain both positive and negative categories, one color can be associated with different emotional tendencies, and certain emotions may be associated with different colors [22].

Color feature extraction typically uses RGB and HSV color spaces. From a human visual perception perspective, the HSV system is closer to psychological responses and color perception, thus more aligned with human observation of color attributes and has visual consistency [23].

The global histogram is most widely used for color feature extraction, representing an image's color distribution [25]. The global color histogram describes the proportion of different colors in an image and is essentially a 1-D discrete function defined as:

$$H(k) = \frac{n_k}{N}, \quad k = 1, 2, 3, \dots, L-1 \quad (1)$$

where k represents the grayscale value, L is the number of possible grayscale values, n_k is the total number of pixels with grayscale value k, and N is the total number of pixels.

The color histogram's advantages are simple calculation—only requiring pixel frequency computation for each color—and strong stability against translation, rotation, or scaling [26]. Researchers primarily use Matlab for color feature extraction.

[Figure 1: see original paper] shows image feature extraction based on color histograms: first processing H, S, and V feature sets separately to obtain small color intervals, then generating histogram bins for corresponding intervals, quantifying the color histogram by calculating color distribution in each interval. [Figure 2: see original paper] shows the quantized color space dimension table, obtaining image color features. The disadvantage is inability to accurately distinguish images with similar color distributions—red and green quantization results appear similar, causing confusion between images with different emotions but similar colors. Additionally, global histograms cannot consider regional features [27].

Texture features also affect human psychology. Though less obvious than color, texture contains contrast and spatial frequency factors that produce different visual effects and emotional impacts. Texture feature extraction primarily uses statistical and geometric methods, with the Gray-Level Co-occurrence Matrix (GLCM) being common.

Traditional algorithms can extract low-level semantic features well but cannot accurately express image emotions, requiring subjective human judgment to establish emotional relationships. The extracted emotional tendencies are highly subjective and cannot objectively represent high-level emotional semantic features.

3. Image Emotion Extraction Based on Improved CNN Models

Deep learning, emerging from artificial neural network research, is a new branch of machine learning. Its core lies in using algorithms to enable computers to simulate human brain processing for recognizing images, audio, and video, ultimately acquiring human-like processing capabilities through training. Common models include Deep Autoencoder (DAE), Deep Restricted Boltzmann Machine (DRBM), and Convolutional Neural Networks (CNN), with CNN achieving significant success in image and face recognition [28].

3.1 CNN Algorithm and Its Application in Image Emotion Extraction

Convolutional Neural Networks (CNN) have proven capable of learning complex mapping relationships from large data series, achieving significant results in computer vision for image classification, object detection, expression recognition, and face detection. Compared with color and texture extraction, deep CNN has strong anti-interference capabilities and high insensitivity to image movement, rotation, deformation, or other transformations [28]. The network structure is shown in [Figure 3: see original paper].

In CNN, the visible layer input is each image patch. In convolutional layers, local features are extracted through filters and nonlinear transformations. The input image is convolved with three trainable filters and additive biases, producing feature maps in layer C1. After pooling in layer S2, these feature maps undergo filtering to obtain layer C3, repeating the calculation process to produce S4. Finally, pixel values are rasterized and connected into a vector input to a traditional neural network for output [28]. The calculation process is:

$$X_j^l = f\left(\sum_{i\in M_j} X_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (2)$$

where M_j is the set of input feature maps, l represents the current convolutional layer, X_j^l is the output feature vector of the jth neuron in layer S, X_i^{l-1} is the output of the ith neuron in the previous convolutional layer, and k_{ij}^l is the connection threshold.

To reduce training parameters, the model introduces parameter sharing: if a feature detector (e.g., vertical edge detection) applies to one image region, it likely applies to others. Each feature detector and output can use the same parameters across different input regions.

Feature vectors from convolutional layers are typically high-dimensional with redundant information, increasing training costs and causing overfitting. Therefore, downsampling (pooling) is performed in S layers to reduce feature map height and width, decreasing computational load and stabilizing feature detector positions. Two pooling types exist:

(1) **Max pooling**: Sliding an (f, f) window across input and storing maximum values. As shown in [Figure 4: see original paper], using a 2$×$2 filter with stride=2 reduces parameters to 1/4 of the original.

(2) **Average pooling**: Sliding an (f, f) window and storing average values. As shown in [Figure 5: see original paper], using a 2$×$2 filter with stride=2 better preserves background features.

Deep networks combining multiple convolutional and pooling layers achieve scale, translation, and rotation invariance. Training uses backpropagation to update network parameters:

$$dA = \sum_{h=0}^{n_H} \sum_{w=0}^{n_H} W_c \times dZ_{hw} \quad (4)$$

where W_c is the filter and dZ_{hw} is the scalar gradient corresponding to the convolutional layer output at row h, column w.

The loss function is introduced in the final layer to represent detection errors, with filter derivatives:

$$dW_c = \sum_{h=0}^{n_H} \sum_{w=0}^{n_H} a_{slice} \times dZ_{hw} \quad (5)$$

where a_{slice} corresponds to the slice producing activation Z_{ij}.

3.2 CNN Model Improvement

The original VGGNet-16 model [29] has obvious defects: (1) Six convolutional sections generate too many parameters, often exceeding available training samples and hindering convergence; (2) Five max pooling layers cause feature vector fragmentation and principal component loss, failing to effectively address overfitting.

Based on pre-training results showing low accuracy and obvious overfitting, this experiment improved the VGGNet model by: (1) Using 5 convolutional layers (conv2d) to optimize for image invariance and prevent overfitting; (2) Setting 4 max pooling layers (maxpool); (3) Using 3 fully connected layers (Dense); (4) Applying dropout to reduce overfitting. Improved model parameters are shown in .

The improved model completes emotion feature extraction with fewer parameters and iterations, with deeper convolutional layers showing higher sensitivity to emotional features. Four pooling layers all use max pooling for dimensionality reduction, effectively increasing feature saliency and providing generalization capability.

The improved CNN algorithm code is described as:

```
# Create convolutional layer and store parameters
def conv_{op}(input_{op}, name, kh, kw, n_{out}, dh, dw, p):
    conv = tf.nn.conv2d(input_{op}, kernel, (1, dh, dw, 1), padding='SAME')
    activation = tf.nn.relu(conv, name=scope)
    return activation

# Define fully connected layer
def fc_{op}(input_{op}, name, n_{out}, p):
    activation = tf.nn.relu_{layer}(input_{op}, kernel, biases, name=scope)
    return activation
```

```python
# Define max pooling layer
def mpool_{op}(input_{op}, name, kh, kw, dh, dw):
    return max_{pool}(input_{op}, ksize=[1, kh, kw, 1], strides=[1, dh, dw, 1],
                    padding='SAME', name=name)

# Define network structure
def inference_{op}(input_{op}, keep_{prob}):
    conv1_1 = conv_{op}(input_{op}, name='conv1_1', kh=3, kw=3, n_{out}=32, dh=1, dw=1, p=p)
    pool1 = mpool_{op}(conv1_2, name='pool1', kh=2, kw=2, dw=2, dh=2)
    # ... (additional layers)
    fc8 = fc_{op}(fc7_{drop}, name="fc8", n_{out}=2, p=p)
    softmax = tf.nn.softmax(fc8)
    predictions = tf.argmax(softmax, 1)
    return predictions, softmax, fc8, p
```

3.3 Building the Image Emotion Extraction Experimental Model

The experimental process requires training a CNN model for image emotion semantics. Different models have different characteristics. After pre-training, testing, and feedback, the CNN-VGGNet model was selected for its good performance. The improved CNN network model is shown in [Figure 7: see original paper], with all convolutional layers configured identically: 3$\times 3 kernels, stride = 1, padding = 1; four 2\times$2 max pooling layers with stride=2; three fully connected layers (first two with 4096 channels, third with 1000 label categories); and ReLU activation in all hidden layers.

[Figure 6: see original paper] shows the improved VGGNet model with 16 total layers (13 convolutional, 3 fully connected). The first convolutional layer transforms input to 32$\times 32\times 64, applying 3\times 3 filters across 3 channels, followed by ReLU, normalization, and max poo
features. The 14th layer is fully connected with 4096 nodes, followed by ReLU, dropout, and final softmax output with fused labels.

4. Experimental Analysis and Evaluation

4.1 Data Source

This study uses a small-sample deep learning approach. The FlickrEmotion dataset [30] was selected, containing 5,000 high-quality social media images with emotional labels. Images are categorized into two emotions: positive (happy, joyful, pleasant) and negative (sad, depressed, suppressed), forming a binary classification problem validated by Softmax. The dataset was split into 2,500 training and 2,500 test samples, with examples shown in [Figure 7: see original paper] and [Figure 8: see original paper].

4.2 Data Preprocessing

To better capture features, images underwent batch preprocessing. Since color and texture effectively express emotion, preprocessing adjusted their weights before CNN input. [Figure 9: see original paper] shows texture enhancement with reduced color features, while [Figure 10: see original paper] shows color enhancement with reduced texture.

Pre-training results in show both preprocessing types achieve ~78% accuracy, indicating good performance but with retrieval loss and overfitting beginning to appear. The improved VGGNet demonstrates good robustness, requiring only fine-tuning for better accuracy.

4.3 Feature Extraction Experiments and Analysis

Preprocessed images were input to the CNN model. With learning rate=0.01 and $3\$\times 3 filters, initialemotionextractionwasperformed.Toimproveaccuracy: (1) Fortexture-enhancednetworks, filtersizewasadjustedto 5\times 5; (2) Forcolor-enhancednetworks, filtersizeremained 3\times\$3$. Adding more convolutional layers caused overfitting; reducing learning rate prevented convergence; increasing it caused neuron death. The final learning rate was set to 0.1.

Trained models extracted emotional features: results approaching 1 indicate stronger positive emotion, while lower values indicate negative emotion. [Figure 11: see original paper] shows a color-enhanced sample classified as 85.22% positive, matching its label. [Figure 12: see original paper] shows a texture-enhanced sample classified as 84.32% positive.

The improved CNN model achieved 10 percentage points higher accuracy than the original VGGNet16. Performance validation on the dataset ([Figure 13: see original paper] and [Figure 14: see original paper]) shows:

- **Texture-enhanced**: Test accuracy peaked at 65% around iterations 20-30, stabilizing near 57% after iteration 50 with clear overfitting as training and test accuracies diverged.
- **Color-enhanced**: Test accuracy peaked at 76% around iterations 60-79, stabilizing near 65% after iteration 80 with less overfitting.

compares the two approaches: texture enhancement achieved 65.35% peak accuracy with 80.21% training accuracy, while color enhancement achieved 75.2% peak accuracy with 77.4% training accuracy. Color features demonstrate better perceptual power for emotion extraction.

Conclusion

This study explores image emotional semantic feature extraction using an improved VGGNet model, achieving 80.21% precision on the dataset—10% higher than the original model. Optimizing network parameters effectively extracts emotional features with satisfactory accuracy. This method eliminates manual intervention, utilizing image properties for emotion determination, and shows strong adaptability for applications like facial emotion recognition [31] and driver safety assessment.

Cross-modal retrieval represents the future of information retrieval, with image emotion-based retrieval being a cutting-edge topic [32-34]. While limited by platform constraints to 4096-dimensional feature vectors, deeper networks could achieve better results.

References

[1] Zeng Jin, Lu Na, Hu Xiaoge, et al. Research on emotional tendency of website news figures [J]. Information Science, 2018, 36(6): 131-135, 141.
[2] Ma Songyue, Xu Xin. Research on user online evaluation based on comment sentiment analysis: A case study of Douban movies [J]. Library and Information Service, 2016, 60(10): 95-102.
[3] Jiang Zhiyi, Ma Wangrong, Zou Kai, et al. Research on emotional evolution characteristics of online public opinion based on sentiment tendency analysis [J]. Modern Intelligence, 2018, 38(4): 50-57.
[4] Huang ?, Lai Maosheng. Research on image retrieval based on user emotion [J]. Information Science, 2006(09): 1395-1399.
[5] Tang Liping, Chen Fen. Research progress on emotion-based image classification [J]. Information Theory and Practice, 2018, 41(6): 149-153, 160.
[6] Li Zhiyi, Huang Zifeng, Xu Xiaomian. Review of cross-modal retrieval models and feature extraction based on representation learning [J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(4): 422-435.
[7] Zeng Jin, Lu Wei, Chen Haihua, et al. Research on microblog user interest identification based on multi-modal data [J]. Information Science, 2018, 36(1): 124-129.
[8] BLACK JA, KUCHI P, PANCHANATHAN S. Indexing natural images for retrieval based on Kansei factors [C]//Proceedings of SPIE - The International Society for Optical Engineering, 2004, 5292: 363-375.
[9] YOSHIDA K, KATO T, YANARU T. Image retrieval system using impression words [C]//IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 1998: 2780-2784 vol.3.
[10] CHO SB, LEE JY. A human-oriented image retrieval system using interactive genetic algorithm [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A, 2002, 32(3): 452-458.
[11] COLOMBO C, DEL BIMBO A, PALA P. Semantic visual information retrieval [J]. IEEE Multimedia, 1999, 6(3): 38-53.
[12] SIERSDORFER S, MINACK E, DENG F, et al. Analyzing and predicting sentiment of images on the social web [C]//Proceedings of the International Conference on Multimedia. ACM Press, 2010: 715-718.
[13] YOU Q, LUO J, JIN H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks [C]//Proceedings of AAAI Conference on Artificial Intelligence, 2015: 381-388.

[14] ORIOL V, ALEXANDER T, SAMY B, et al. Show and tell: A neural image caption generator [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3156-3164.
[15] SUN M, SONG Z, JIANG X, et al. Learning pooling for convolutional neural network [J]. Neurocomputing, 2017, 224: 96-104.
[16] Huang ?, Luo Fang, You ?. Image emotional features and retrieval applications [J]. Information Science, 2010, 28(04): 602-606.
[17] Wang Shangfei, Xue Jia, Wang Xufa. Content-based affective image acquisition model [J]. Computer Science, 2004(9): 186-189.
[18] Wu Pin, Tao Cong, Zhu Yonghua, et al. Image annotation and retrieval based on emotional semantics [J]. Computer Technology and Development, 2015, 25(10): 13-18.
[19] Wang Huaqiu, Hu Lisong. Research on image emotional semantic classification and retrieval [J]. Journal of Chongqing University of Technology (Natural Science), 2017, 31(10): 180-186.
[20] Tan Lianzhi. Research on face attribute recognition based on deep learning [D]. Beijing: University of Chinese Academy of Sciences (Shenzhen Institutes of Advanced Technology), 2017.
[21] Chen Jinju, Ou Shiyan. Comparison and analysis of digital image semantic annotation models [J]. Library and Information Service, 2014, 37(5): 92-96, 106.
[22] Chen Jing, Chen Dezhao, Lu Quan. An image user interest extraction model based on expression analysis [J]. Information Theory and Practice, 2014, 37(5): 92-96, 106.
[23] Jiang Yaxi. Image color editing considering emotional factors [D]. Tianjin: Tianjin University, 2017.
[24] Xu Tongyang, Ren Haoran, Zhang Guobiao. Research on frontier dynamics of image retrieval in China: Bibliometric analysis of CNKI data [J]. Library Science Research, 2017(4): 2-10.
[25] Cheng Lin, Chen Junjie, Xiang Jie. Research and application of image color feature extraction technology [J]. Computer Engineering and Design, 2009, 30(14): 3451-3454.
[26] Wang Yanlin. Implementation of HSV color feature-based image retrieval algorithm in Matlab [J]. Computer Programming Skills & Maintenance, 2013(16): 86-87.
[27] Yang Rui. Design and implementation of retrieval system based on color emotional semantics [D]. Kaifeng: Henan University, 2015.
[28] Zhang Junyang, Wang Huili, Guo Yang, et al. Review of deep learning research [J]. Computer Application Research, 2018, 35(7): 1921-1928, 1936.
[29] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014. arXiv:1409.1556.
[30] Lu Wei, Luo Mengqi, Ding Heng, et al. Comparative study of deep learning image annotation and user annotation [J]. Data Analysis and Knowledge Discovery, 2018, 2(5): 1-10.
[31] Zhu Hong, Li Qianmu, Li Deqiang. Facial multi-feature point localization based on single convolutional neural network [J]. Computer Science, 2018, 45(4):

273-277, 284.
[32] Hu Erlei, Feng Rui. Image retrieval system based on deep learning [J]. Computer Systems & Applications, 2017, 26(3): 8-19.
[33] Ren Xiali, Chen Guangxi, Cao Jianshou, et al. Image retrieval method based on deep learning features [J]. Computer Engineering and Design, 2018, 39(2): 503-510.
[34] Wu Qiong. Research on image-text sentiment analysis technology based on convolutional neural networks [D]. Quanzhou: Huaqiao University, 2017.
[35] Jin Hanjun, Duan Beibei. Research on application of convolutional neural networks in cross-media retrieval [J]. Electronic Measurement Technology, 2018, 41(7): 54-57.

Author Contributions

Li Zhiyi: Conceptualization, research design, algorithm coding, paper revision and finalization.
Xu Hongkai: Paper drafting, algorithm improvement, testing, reference analysis.
Duan Bin: Platform construction, algorithm improvement, parameter setting.

Research on Image Emotion Feature Extraction Based on Deep Learning CNN Model

Li Zhiyi[1], Xu Hongkai[1], Duan Bin[2]
[1]School of Economics & Management, South China Normal University, Guangzhou 510006
[2]School of Information & Optoelectronic Science, South China Normal University, Guangzhou 510006

Abstract: [Purpose/significance] Image retrieval based on user emotion has become a hotspot in machine learning research. However, corpus data for image sentiment feature annotation mostly derives from low-level feature extraction, leading to simplified and stylized retrieval processes. This paper proposes a deep learning-based algorithm that fuses low-level features with high-level emotional semantics, providing reference for emotional semantic image retrieval. [Method/process] Using an improved CNN model, color and texture features are input as dataset images, emotional information is automatically extracted through multi-layer operations, and sentiment retrieval accuracy is calculated via backpropagation to construct a high-accuracy, low-overfitting extraction model. [Result/conclusion] The improved CNN model achieves image emotion feature extraction, improving retrieval accuracy by 10% compared to the original model.

Keywords: deep learning; image; emotional features; extraction; convolutional neural network

---

Minutes of the Inaugural Meeting of the 2018-2022 Ministry of Education Library Science Teaching Guidance Committee

On January 18, 2019, the inaugural meeting of the 2018-2022 Ministry of Education Library Science Teaching Guidance Committee (hereinafter "the Committee") was held at Shanghai University. Committee Chair Professor Wang Yuguang (Peking University), Vice Chairs Professor Chen Chuanfu (Wuhan University), Professor Zheng Jianming (Nanjing University), Professor Xia Lixin (Central China Normal University), and Professor Lü Bin (Shanghai University), Secretary-General Professor Zhang Jiuzhen (Peking University), and over 30 committee members from domestic library science programs gathered at Shanghai University to celebrate the Committee's establishment and plan future work.

Shanghai University Vice President Nie Qing attended and delivered opening remarks, introducing Shanghai University's high-level university initiatives and undergraduate education measures. Chair Wang Yuguang expressed gratitude to Shanghai University, conveyed the Ministry of Education's guidance spirit, and presented appointment letters to committee members.

The first working session, chaired by Vice Chair Xia Lixin, featured a work summary report by Vice Chair Chen Chuanfu reviewing the Committee's development and achievements. All members studied the Ministry of Education's Committee Charter and the Library Science Undergraduate Education National Standards. Secretary-General Zhang Jiuzhen conveyed the Ministry's "undergraduate-first" and "comprehensive revitalization of undergraduate education" requirements, presenting the five-year work plan covering national standard implementation, teaching quality improvement, curriculum development, textbook construction, talent cultivation research, and faculty development, which was approved.

The discussion segment, chaired by Vice Chair Lü Bin, featured suggestions from various members. Chair Wang Yuguang concluded by summarizing the meeting and arranging the second 2019 meeting. The meeting concluded successfully after completing all agenda items.

(Committee member list appears on page 126)

Note: Figure translations are in progress. See original paper for figures.