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

Traitement du Signal
Vol. 39, No. 3, June, 2022, pp. 931-937
Journal homepage: http://iieta.org/journals/ts

A Deep Learning-Based Cluster Analysis Method for Large-Scale Multi-Label Images

Yanping Xu

School of Information Engineering, Yancheng Institute of Technology, Yancheng 224051, China

Corresponding Author Email: xuyp@ycit.cn

https://doi.org/10.18280/ts.390319

Received: 9 January 2022
Accepted: 17 April 2022

Keywords:
deep learning, large-scale, multi-label images, cluster analysis

ABSTRACT

Large-scale multi-label image classification requires determining the presence or absence of a target object in a large number of sample images. For highly specialized and complex multi-label image sets, it is especially important to ensure the accuracy of image classification. Traditional deep learning models usually don't take into account image-label correlation constraints when classifying multi-label images, and the strategy of classifying images based only on their own features greatly limits the model performance. In this context, this paper focuses a deep learning-based cluster analysis method for large-scale multi-label images. We constructed a model for large-scale multi-label image category recognition, which consists of a global image feature extraction module, a feature activation vector generation module and an image category inter-label connection module. Using a graph convolutional network (GCN), we aggregated the information of image category label nodes in the constructed multi-label graph structure, while exploring the correlation between image category labels. A detailed description is presented on how to introduce the attention mechanism into the constructed model mentioned above for image category recognition. Experimental results have validated the effectiveness of the constructed model.

1. INTRODUCTION

As an important branch in the field of computer vision, image classification has been widely used in target recognition, defect detection and other application scenarios [1-8]. By the number of tags in the image, there are two types of image classification: single-label image classification and multi-label image classification. With the development and promotion of deep learning technology, the performance of single-label image classification methods is already superior enough. Compared to single-label image classification, multi-label image classification responds to a more common demand. Its implementation and realization are more difficult and complex, hence more challenging in image processing tasks [9-15]. Large-scale multi-label image classification requires the determination of the presence or absence of a target object in a large number of sample images [16-24]. Due to the huge quantity of samples and the varying awareness levels of personnel, this could lead to very low efficiency in multi-label image classification, especially for highly specialised and complex multi-label image sets. Therefore, it is particularly important to ensure the accuracy of image classification.

Existing multi-label image classification methods focus on the accuracy of label prediction and ignore the structural information embedded in the hierarchical label space. To address these issues, Wang et al. [25] proposed a hierarchical framework based on the feature and label structural information named Hierarchical GAN-Tree and Bi-Directional Capsules (HGT&BC), which generates hierarchical feature space using the unsupervised divisive clustering pattern, alleviating the mode-collapse of generators and the overfitting manifestation of conventional GANs. Traditional approaches use attention mechanisms or prior knowledge but lack deep semantic associations, resulting in degraded detection performance. Yao et al. [26] proposed a brain-inspired memory graph convolutional network (M-GCN). M-GCN presents crucial short-term and long-term memory modules to interact attention and prior knowledge, learning complex semantic enhancement, and suppression. Extensive experiments demonstrate that M-GCN outperforms general state-of-the-art methods and shows the advantages in semantic correlation and complexity comparing with traditional memory models. In multi-label image retrieval, existing deep hashing simply indicates whether two images are similar by constructing a similarity matrix. To fulfil this gap, Shen et al. [27] proposed Deep Co-Image-Label Hashing (DCILH) to discover label dependency. Specifically, DCILH regards image and label as two views, and maps the two views into a common deep Hamming space. To exploit label dependency, DCILH further employs the label-correlation aware loss on the predicted labels, such that predicted output on positive label is enforced to be larger than that on negative label. Wang et al. [28] developed Cross-modal Fusion for Multi-label Image Classification with attention mechanism (termed as CFMIC), which combines attention mechanism and GCN to capture the local and global label dependencies simultaneously in an end-to-end manner. Extensive experiments on MS-COCO and VOC2007 verified CFMIC greatly promotes the convergence efficiency and produces better classification results than the state-of-the-art approaches. Wang et al. [29] proposed a multi-attention fusion network with dilated convolution and label smoothing for content-based remote sensing image retrieval (CBRSIR). First, a dilated convolutional layer was used to replace the fifth convolutional layer in the network to obtain a large receptive field. Besides, in order to enhance the differences between the

discriminative features of those correct and incorrect classes, label smoothing was used to replace the cross-entropy loss function. Experimental results illustrated that such network can be effectively migrated to other similar convolutional neural network (CNN) models and can achieve state-of-the-art or competitive results.

Traditional deep learning models usually don't take into account image-label correlation constraints when classifying multi-label images, and the strategy of classifying images based only on their own features greatly limits the model performance. How to fully exploit the label co-occurrence in sample images and ensure that the model's classification performance is sufficiently satisfactory remains a research topic of much scholarly attention currently. In this context, this paper focuses a deep learning-based cluster analysis method for large-scale multi-label images. The paper unfolds the following major aspects: (1) We constructed a model for large-scale multi-label image category recognition, which consists of a global image feature extraction module, a feature activation vector generation module and an image category inter-label connection module. (2) Using a graph convolutional network (GCN), we aggregated the information of image category label nodes in the constructed multi-label graph structure, while exploring the correlation between image category labels. (3) A detailed description is presented on how to introduce the attention mechanism into the constructed model mentioned above for image category recognition. Experimental results have validated the effectiveness of the constructed model.

2. CONSTRUCTION OF A MODEL FOR LARGE-SCALE MULTI-LABEL IMAGE CATEGORY RECOGNITION

In this paper, we construct a model for large-scale multi-label image category recognition, which consists of a global image feature extraction module, a feature activation vector generation module and an image category inter-label connection module. In the global image feature extraction module, the image features are extracted by *ResNet*-101 CNN. Assume that the input image definition is represented by *RT*, the parameters that can be learned by the global image feature extraction module are represented by ω_{dmm}, the global average pooling (GAP) operation is represented by g_{pool} (.), and the dimensionality of the global image features is denoted by *E*. Then the global image feature *a* after processing by the global image feature extraction module and the GAP operation can be expressed as:

$$a = g_{pool}\left(g_{dmm}\left(RT;\omega_{dmm}\right)\right) \in S^{E} \tag{1}$$

Assume that the parameters of the fully connected layer are denoted by $\omega_{gd} \in R^{D\times D}$ and the number of sample image categories is denoted by *D*. Inputting the global feature *a* into the fully connected layer of the module yields an initial multi-label image category recognition result b_{cls}:

$$b_{cls} = g_{gd}\left(a;\omega_{gd}\right) \tag{2}$$

To construct category label relationships between images, the feature activation vector generation module decouples the global features extracted by the global image feature extraction module to generate the feature activation vectors corresponding to different image categories. Assume that the feature activation vectors corresponding to different image categories are represented by *C*. The process of copying $a \in R^{E}$ *D* times to obtain $[a,...,a]^{T} \in S^{D\times E}$" is denoted by $g_{copy}(a)$. The Hadamard product is denoted by $\oplus$. The following can be obtained by filtering the obtained global feature *a* based on ω_{gd}:

$$C = g_{copy}\left(a\right)\oplus\omega_{gd} \in S^{D\times D} \tag{3}$$

After obtaining the feature activation vectors corresponding to different image categories, this paper constructs the adjacency matrix between image category labels based on GCN, so as to extract graph structure information from multi-label image sample sets.

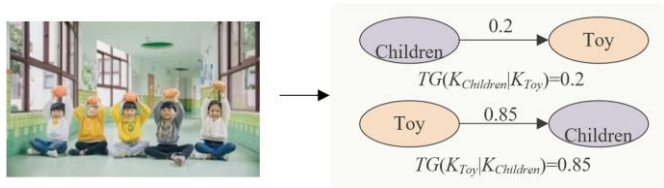

Figure 1. Schematic diagram of the probability of a category label condition

Defining the relationship between image category labels as a conditional probability, we kept statistics on the number of simultaneous occurrences of two image category labels in the image sample set, and constructed a frequency matrix using the statistics as elements, which is denoted by $N \in S^{D\times F}$. Figure 1 gives a schematic diagram of the conditional probabilities of category labels. Suppose N_{ij} denotes the number of simultaneous occurrences of image category labels *i* and *j*, M_i denotes the number of occurrences of category label *i* in the image sample set, and $TG_{ij}=TG(K_j|K_i)$ denotes the conditional probability of category label *j* occurring when category label *i* occurs. Then the conditional probability matrix characterizing the relationship between image category labels can be obtained based on *N* as follows:

$$TG_i = N_i / M_i \tag{4}$$

Using the conditional probability matrix shown in equation 4 directly as an adjacency matrix between image category labels can result in a large amount of noise in the co-occurrence patterns between labels or in matrix overfitting. The solution is to binarise equation 4. Assuming that the mean of the probability matrix *TG is* represented by *AV*(*TG*) and the standard deviation is represented by *SD*(*TG*), the binarisation process is as follows:

$$X_{ij} = \begin{cases} 0, if \quad TG_{ij} < AV\left(TG\right) - SD\left(TG\right) \\ 1, if \quad TG_{ij} \geq AV\left(TG\right) - SD\left(TG\right) \end{cases} \tag{5}$$

In order to construct relationships between image category labels, further multi-label graph structures can be generated based on the initial multi-label image category recognition results. After feeding *C* and *TG* into the GCN, we obtained the output auxiliary multi-label image category recognition result, which is denoted by b_{gcn}. Assuming the GCN parameters are denoted by ω_{gcn}, we had

$$b_{gcn} = g_{GCN}\left(C, X; \omega_{gcn}\right) \tag{6}$$

The final multi-label image category recognition result, represented by b^*, is obtained by overlaying b_{cls} with b_{gcn}

$$b^* = b_{cls} + b_{gcn} \tag{7}$$

If the result satisfies $b \in S^D$, it is assumed that whether category label i appears in the image is characterized by b^i with values 0, 1. Assuming that the sigmoid function is represented by $\varepsilon(\cdot)$, the loss function of the constructed model for large-scale multi-label image category recognition can be represented by the following equation:

$$K = -d = 1\sum_{d=1}^{D} b^d log\left(\varepsilon\left(b^{d*}\right)\right) + \left(1-b^d\right) log\left(1-\varepsilon\left(b^{d*}\right)\right) \tag{8}$$

3. INTRODUCTION OF INFORMATION FUSION MODULE

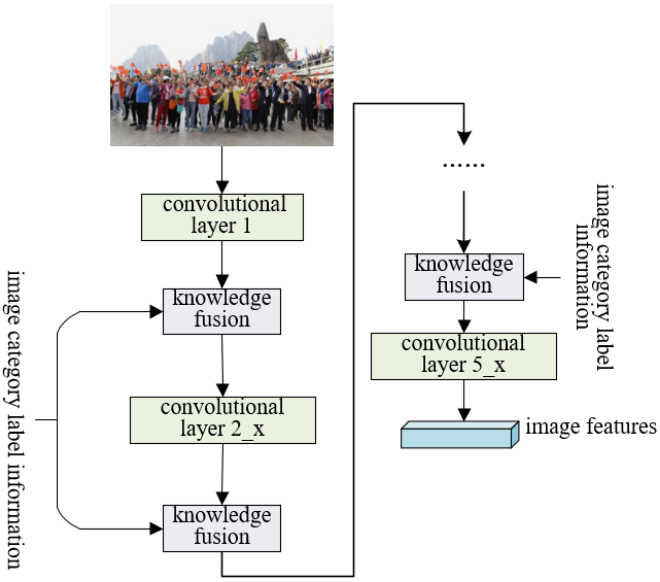

Figure 2. Information fusion process of image category label nodes

In order to explore the correlation between image category labels, information aggregation of image category label nodes is required using a GCN in the constructed multi-label graph structure. Figure 2 presents the information fusion process of the image category label nodes. The process of information aggregation is expressed through Eq. (9) as:

$$P^{k+1} = \varepsilon\left(X' P^k Q^k\right) \tag{9}$$

It is assumed that X' denotes the adjacency matrix after normalization, $P^k \in S^{M \times Dk}$ denotes the feature matrix of the GCN at the k-th layer containing information on all image category label nodes, and $Q^k \in S^{Dk \times Dk+1}$ denotes the parameter matrix of the completed training network.

For traditional multi-label image recognition methods, association rule mining of image categories is usually achieved through classifiers, while global image features are extracted through neural networks. There is minimal correlation between the two behaviours. To solve this problem, this paper introduces the idea of knowledge fusion into the constructed model for large-scale multi-label image category recognition, i.e., the association between global image features is completed during the image feature extraction stage.

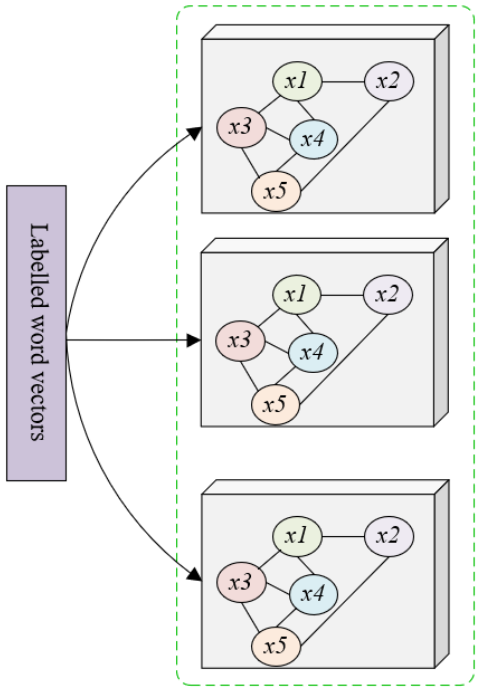

Figure 3. Image category label feature information extraction module

Assume that $A \in S^{D \times F \times Q}$ denotes the global image features extracted by the convolution operation shown in Figure 3, D denotes the number of channels of the global image features, while F and Q denote the length and width of the image feature map respectively, $P \in S^{M \times D}$ denotes the GCN output feature information, M denotes the total number of category labels, and $g(T)$ denotes the convolution operation to achieve fusion of the category label relationship information into the global image features. We can accomplish the dimension transformation of the category label relationship feature from $A \in S^{M \times F \times Q}$ to $S^{D \times F \times Q}$. If the Tanh activation function is denoted by $\varepsilon(.)$, and the operations on the array shape transformations are denoted by $O_{(M \times F \times Q)}(.)$ and $O_{(FQ \times D)}$, then the introduced knowledge fusion module expression is given by the following equation:

$$B = g\left(O_{(M \times F \times Q)}\left(O_{(FQ \times D)}\left(A\right)\varepsilon\left(P^O\right)\right)\right) + A \tag{10}$$

Applying $Q^{\wedge} \in S^{D \times E}$ to the i-th global image feature $a_i \in S^E$, the multi-label image category recognition result is obtained as:

$$\hat{b} = \hat{Q} a_i \tag{11}$$

The corresponding loss function of the model can be expressed by the following equation:

$$K_{cls} = -\frac{1}{D}\sum_{j=1}^{D} b_i^j log\left(\varepsilon\left(\hat{b}_i^j\right)\right) + \left(1-b_i^j\right) log\left(1-\varepsilon\left(\hat{b}_i^j\right)\right) \tag{12}$$

4. INTRODUCTION OF ATTENTION MECHANISM

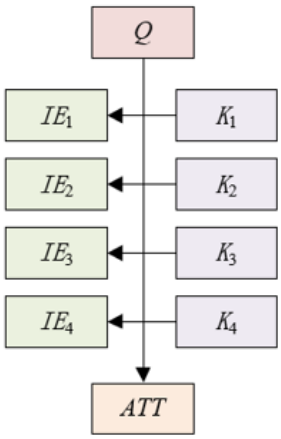

Figure 4. Schematic diagram of attention processes

When exposed to an image, the brain's signal processing mechanism, which is unique to human vision, will suppress some useless information and devote more attention to important targets only. In this paper, the attention mechanism is introduced into the constructed model for large-scale multi-label image category recognition to obtain better results in image category recognition. The attention process is illustrated in Figure 4. The introduced attention mechanism process is divided into three main steps:

First, we used a function to calculate the correlation between "query (Q)" and each "keyword (K)", which can be obtained by vector dot product or by vector similarity. The result is expressed by Ω_i.

Next, Ω_i is normalized based on the softmax function, while the weight x_i corresponding to the significant element IE_i is emphasized.

$$x_i = Softmax(\Omega_i) = \frac{p^{\Omega_i}}{\sum_{j=1}^{K_a} p^{\Omega_j}} \tag{13}$$

Compute the weighted sum of the weighted values calculated in equation 13 to obtain the desired attention value:

$$ATT(Q', S') = \sum_{j=1}^{K_a} x_i \bullet IE_i \tag{14}$$

Below is a detailed description of how the attention mechanism is introduced into the constructed model for large-scale multi-label image category recognition to carry out image category recognition tasks. Figure 5 presents the construction process of label relationship.

Assuming that M denotes the number of targets, i.e., labels, contained the input image, g_X denotes the extracted local image features, and g_H denotes the image label location features, the input image can then be represented as $\{(g^m{}_X, g^m{}_H)\}^M{}_{m=1}$. Suppose that $g^n{}_X$ represents the local image features of the n-th target, Q_u represents the convolution operation with a 1×1 convolution kernel, and q^{nm} represents the relationship weights between different image labels. Equation 15 gives the expression for the relational features between the m-th image label and the other labels:

$$g_S(m) = \sum_n q^{nm} \bullet \left(Q_U \bullet g_X^n\right) \tag{15}$$

The core of the execution of attention mechanism is the fusion of the $g_S(m)$ calculated in the above equation with the g_X as input to the lower-level network for further information transfer. The relational weights q^{nm} between the labels are generated based on the softmax function. Assuming that $Q^{nm}{}_X$ is the feature weight of an image label and $Q^{nm}{}_H$ the position weight, we have

$$q^{nm} = \frac{Q_H^{nm} \cdot exp\left(Q_X^{nm}\right)}{\sum_l Q_H^{lm} \cdot exp\left(Q_X^{lm}\right)} \tag{16}$$

$Q^{nm}{}_X$ and $Q^{nm}{}_H$ can be calculated by Eqns. (17) and (18), where $Q^{nm}{}_X$ is implemented based on the fully connected layer of the network. It is assumed that the parameters of the fully connected layer are represented by Q_L and Q_W respectively, and the dot product operation is represented by $\Psi(.)$.

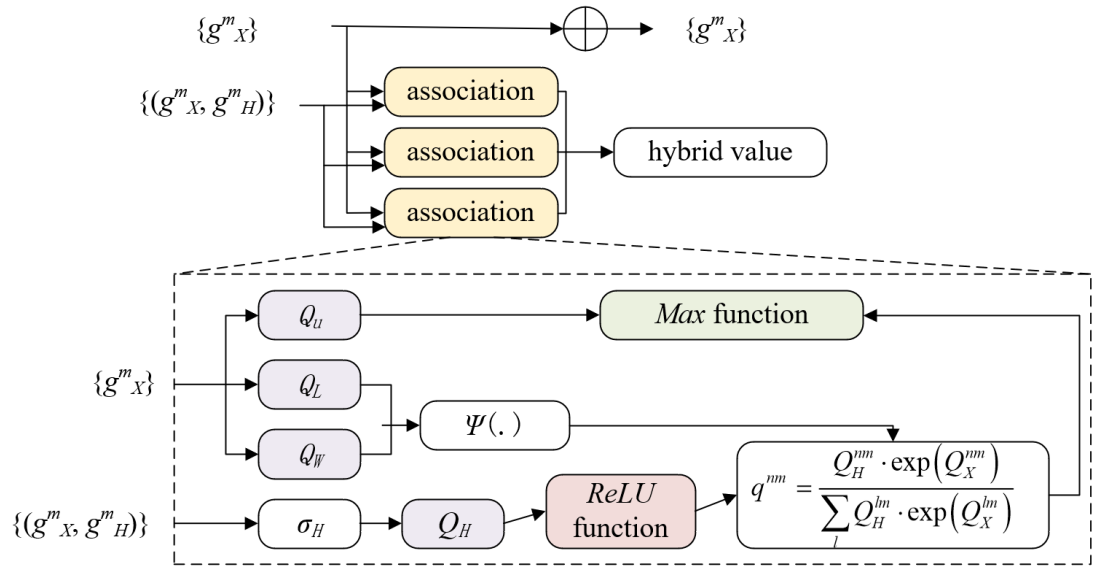

Figure 5. Label relationship construction process

$$q_X^{nm} = \frac{\Psi\left(Q_L g_X^n, Q_W g_X^m\right)}{\sqrt{e_l}} \tag{17}$$

$$q_H^{nm} = max\left\{0, Q_H \cdot \sigma_H\left(g_H^n, g_H^m\right)\right\} \tag{18}$$

The σ_H function in the above equation consists of cosine and sine functions, which functions to achieve a high-dimensional transformation of the position coordinate information of image labels. The max function used to constrain the position weights can be equated by the ReLU activation function. For the network training scattering problem, it requires the coordinate transformation processing of images, i.e., finishing the scale normalisation and data logarithm processing of images to increase their scale invariance. Suppose the coordinate transformation process is represented by g_H, which can be expressed as:

$$\left(log\left(\frac{|a_n - a_m|}{q_n}\right), log\left(\frac{|b_n - b_m|}{f_n}\right), log\left(\frac{q_m}{q_n}\right), log\left(\frac{f_m}{f_n}\right)\right)^T \tag{19}$$

5. EXPERIMENTAL RESULTS AND ANALYSIS

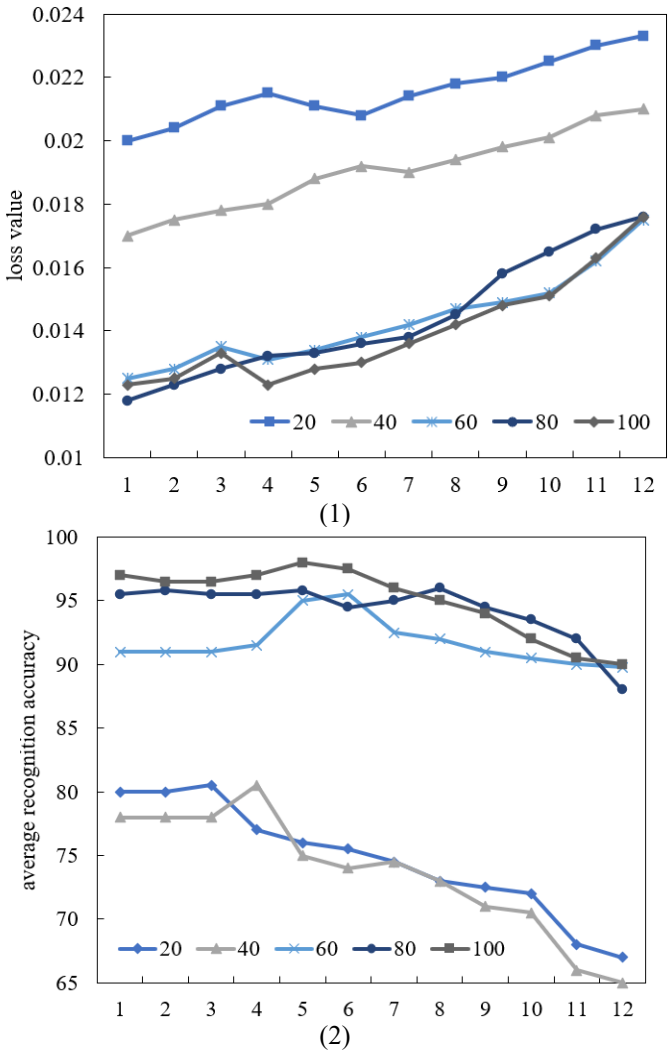

Figure 6. Comparison of metric performance of images in different categories with different number of replications

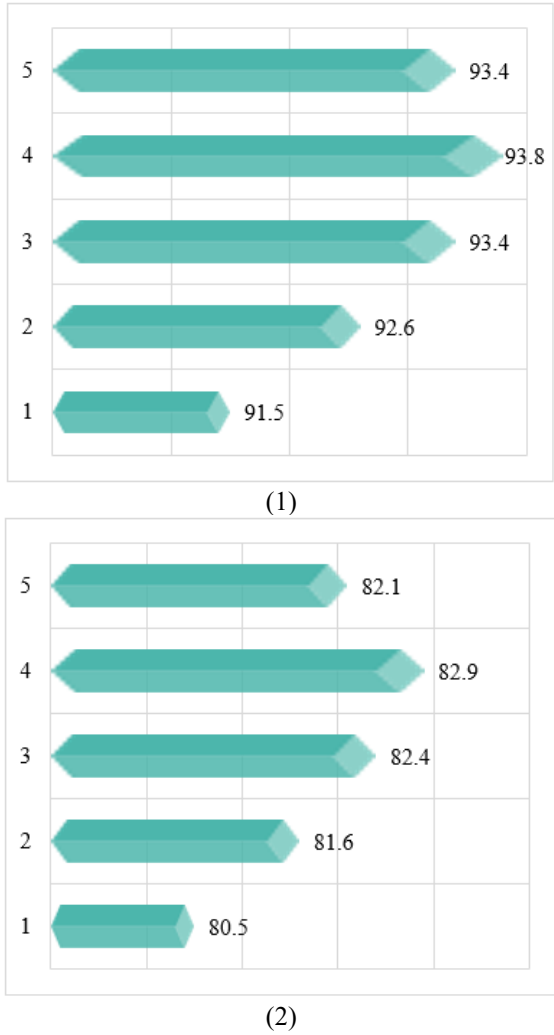

Figure 7. Experimental results of the information fusion module for different data sets

The number of feature replications is a parameter that needs to be set during the generation of the feature activation vectors corresponding to different image classes. Figure 6 gives a performance comparison of the metrics for different categories of images with different number of replications. The performance of the model for different categories of images with different number of replications is evaluated in terms of loss value and average recognition accuracy. The number of replications is increased from 20 to 100 in intervals of 20. As can be seen from the figure, the performance of the model improves significantly as the number of feature replications increases, regardless of image category, until the number of feature replications exceeds 70. Then the performance of the model slowly decreases. Therefore, in this paper, the number of feature replications is set to 70.

When the number of GCN layers is increasing, there are differences in image recognition results. Figure 7 shows the experimental results of the information fusion module for different data sets. It can be seen that when the number of layers is 1 or 2, the network is unable to fully mine and learn the information contained in image labelling relationships. When the number of network layers reaches 5 or more, the information contained in image labelling relationships will be in a state of excessive circulation, along with a significant reduction of information differentiation. The best result is achieved when the number of network layers is 4.

Table 1. Comparison of experimental results of different image category recognition models

Model number	1	2	3	4	5
Data	06+14	06+14	06+14	06+14	06+14
Basic Network	VGG-15	DS/64-192-48-1	ResNet-101	ResNet-101	ResNet-101
Sample size	256 x 256	256 x 256	512 x 512	512 x 512	512 x 512
mAP	72.8	70.1	77.5	81.8	87.2

In this paper, we optimise some existing multi-label image category recognition models, elaborate on optimisation ideas, and verify their effectiveness with reference to experimental results. The model constructed in this paper was trained in the same experimental environment as traditional models. The five models involved in the experiments are Fast-CNN, DHC, ResNet-10, the model before the introduction of the information fusion module, and the model constructed in this paper. Table 1 compares the experimental results of different models. As can be seen from the table, the model constructed in this paper, which integrates the advantages of knowledge fusion and attention mechanisms, outperforms all the other models in terms of accuracy in recognising multi-label image categories. This is due to the fact that the model constructed in this paper completes the association between global image features at the image feature extraction stage and performs coordinate transformation on an image, making full use of the features at each image scale to increase the scale invariance. This model introduces an attention mechanism, which emphasises the more attention-grabbing targets in an image based on image label similarity, ensuring more desirable accuracy in image category recognition.

Next, the experimental results of this model were compared with other classifiers, and the results are shown in Table 2. The three classifiers involved in the experiment are the logistic regression model, the stochastic forest model and the model constructed in this paper. As can be seen from the figure, the accuracy of the model constructed in this paper is higher than that of the other two models, by 5.85% and 10.27% respectively. This demonstrates the effectiveness of this model in identifying and classifying multi-label image categories.

Table 2. Comparison of experimental results of different classifiers

Model number	1	2	3
Data	06+14	06+14	06+14
Basic Network	VGG-15	VGG-15	ResNet-101
mAP	75.1	78.8	83.7

6. CONCLUSION

This paper focuses a deep learning-based cluster analysis method for large-scale multi-label images. We constructed a model for large-scale multi-label image category recognition, which consists of a global image feature extraction module, a feature activation vector generation module and an image category inter-label connection module. Using a graph convolutional network (GCN), we aggregated the information of image category label nodes in the constructed multi-label graph structure, while exploring the correlation between image category labels. A detailed description is presented on how to introduce the attention mechanism into the constructed model mentioned above for image category recognition. We compared the performance metrics for different image categories with different number of replications to determine the number of feature replications. We gave experimental results of the information fusion module for different datasets, and determined the number of GCN layers. The constructed model was trained with conventional models in the same experimental environment. After that, we compared the experimental results of different models to verify the effectiveness of the model in this paper for the recognition and classification of multi-label image categories.

REFERENCES

[1] Tian, C., Sun, G., Zhang, Q., Wang, W., Chen, T., Sun, Y. (2017). Integrating sparse and collaborative representation classifications for image classification. International Journal of Image and Graphics, 17(2): 1750007. https://doi.org/10.1142/S0219467817500073

[2] Singh, A.K., Kim, Y.H. (2021). Classification of drones using edge-enhanced micro-doppler image based on CNN. Traitement du Signal, 38(4): 1033-1039. https://doi.org/10.18280/ts.380413

[3] Liu, C., Li, J., He, L., Plaza, A., Li, S., Li, B. (2020). Naive Gabor networks for hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems, 32(1): 376-390. https://doi.org/10.1109/TNNLS.2020.2978760

[4] Xie, W., Xie, Z., Zhao, F., Ren, B. (2018). POLSAR image classification via clustering-WAE classification model. IEEE Access, 6: 40041-40049. https://doi.org/10.1109/ACCESS.2018.2852768

[5] Yu, C., Song, M., Chang, C. I. (2018). Band subset selection for hyperspectral image classification. Remote Sensing, 10(1): 113. https://doi.org/10.3390/rs10010113

[6] Cao, X.P., Li, T., Bai, J.W., Wei, Z.K. (2020). Identification and classification of surface cracks on concrete members based on image processing. Traitement du Signal, 37(3): 519-525. https://doi.org/10.18280/ts.370320

[7] Kurmi, Y., Gangwar, S., Agrawal, D., Kumar, S., Srivastava, H.S. (2021). Leaf image analysis-based crop diseases classification. Signal, Image and Video Processing, 15(3): 589-597. https://doi.org/10.1007/s11760-020-01780-7

[8] Ji, J., Jiang, L., Zhang, T., Zhong, W., Xiong, H. (2021). Adversarial erasing attention for fine-grained image classification. Multimedia Tools and Applications, 80(15): 22867-22889. https://doi.org/10.1007/s11042-020-08666-3

[9] Tang, C., Liu, X., Wang, P., Zhang, C., Li, M., Wang, L. (2019). Adaptive hypergraph embedded semi-supervised multi-label image annotation. IEEE Transactions on Multimedia, 21(11): 2837-2849. https://doi.org/10.1109/TMM.2019.2909860

[10] Gao, B.B., Zhou, H.Y. (2021). Learning to discover multi-class attentional regions for multi-label image recognition. IEEE Transactions on Image Processing, 30: 5920-5932. https://doi.org/10.1109/TIP.2021.3088605

[11] Ban, X., Li, P., Wang, Q., Zhou, S., Guo, S., Wang, Y. (2021). Graph attention mechanism with global contextual information for multi-label image recognition. Journal of Electronic Imaging, 30(6): 063031. https://doi.org/10.1117/1.JEI.30.6.063031
[12] Chu, W.T., Huang, S.H. (2021). Multi-label image recognition by using semantics consistency, object correlation, and multiple samples. Journal of Visual Communication and Image Representation, 77: 103067. https://doi.org/10.1016/j.jvcir.2021.103067
[13] Wang, L., Zhang, A., Wang, P., Dong, Y. (2019). Automatic image annotation using model fusion and multi-label selection algorithm. Journal of Intelligent & Fuzzy Systems, 37(4): 4999-5008. https://doi.org/10.3233/JIFS-182587
[14] Guo, H.F., Han, L., Su, S., Sun, Z.B. (2018). Deep multi-instance multi-label learning for image annotation. International Journal of Pattern Recognition and Artificial Intelligence, 32(3): 1859005. https://doi.org/10.1142/S021800141859005X
[15] Xia, S., Chen, P., Zhang, J., Li, X., Wang, B. (2017). Utilization of rotation-invariant uniform LBP histogram distribution and statistics of connected regions in automatic image annotation based on multi-label learning. Neurocomputing, 228: 11-18. https://doi.org/10.1016/j.neucom.2016.09.087
[16] Zou, F., Liu, Y., Wang, H., Song, J., Shao, J., Zhou, K., Zheng, S. (2016). Multi-view multi-label learning for image annotation. Multimedia Tools and Applications, 75(20): 12627-12644. https://doi.org/10.1007/s11042-014-2423-2
[17] Kolisnik, B., Hogan, I., Zulkernine, F. (2021). Condition-CNN: A hierarchical multi-label fashion image classification model. Expert Systems with Applications, 182: 115195. https://doi.org/10.1016/j.eswa.2021.115195
[18] Yan, Z., Liu, W., Wen, S., Yang, Y. (2019). Multi-label image classification by feature attention network. IEEE Access, 7: 98005-98013. https://doi.org/10.1109/ACCESS.2019.2929512
[19] Xue, Z., Du, J., Zuo, M., Li, G., Huang, Q. (2019). Label correlation guided deep multi-view image annotation. IEEE Access, 7: 134707-134717. https://doi.org/10.1109/ACCESS.2019.2941542
[20] Lyu, F., Li, L., Victor, S.S., Fu, Q., Hu, F. (2019). Multi-label image classification via coarse-to-fine attention. Chinese Journal of Electronics, 28(6): 1118-1126. https://doi.org/10.1049/cje.2019.07.015
[21] Bao, S., Chung, A. C. (2018). Multi-scale structured CNN with label consistency for brain MR image segmentation. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(1): 113-117. https://doi.org/10.1080/21681163.2016.1182072
[22] Lai, H., Yan, P., Shu, X., Wei, Y., Yan, S. (2016). Instance-aware hashing for multi-label image retrieval. IEEE Transactions on Image Processing, 25(6): 2469-2479. https://doi.org/10.1109/TIP.2016.2545300
[23] Lee, J.H., Rico-Jimenez, J.J., Zhang, C., et al. (2019). Simultaneous label-free autofluorescence and multi-harmonic imaging reveals in vivo structural and metabolic changes in murine skin. Biomedical Optics Express, 10(10): 5431-5444. https://doi.org/10.1364/BOE.10.005431
[24] Liu, L., Plawinski, L., Durrieu, M.C., Audoin, B. (2019). Label-free multi-parametric imaging of single cells: dual picosecond optoacoustic microscopy. Journal of Biophotonics, 12(8): e201900045. https://doi.org/10.1002/jbio.201900045
[25] Wang, B., Hu, X., Zhang, C., Li, P., Philip, S.Y. (2022). Hierarchical GAN-Tree and Bi-Directional Capsules for multi-label image classification. Knowledge-Based Systems, 238: 107882. https://doi.org/10.1016/j.knosys.2021.107882
[26] Yao, X., Xu, F., Gu, M., Wang, P. (2022). M-GCN: Brain-inspired memory graph convolutional network for multi-label image recognition. Neural Computing and Applications, 34(8): 6489-6502. https://doi.org/10.1007/s00521-021-06803-z
[27] Shen, X., Dong, G., Zheng, Y., Lan, L., Tsang, I.W., Sun, Q.S. (2021). Deep co-image-label hashing for multi-label image retrieval. IEEE Transactions on Multimedia, 24: 1116-1126. https://doi.org/10.1109/TMM.2021.3119868
[28] Wang, Y., Xie, Y., Zeng, J., Wang, H., Fan, L., Song, Y. (2022). Cross-modal fusion for multi-label image classification with attention mechanism. Computers and Electrical Engineering, 101: 108002. https://doi.org/10.1016/j.compeleceng.2022.108002
[29] Wang, S., Hou, D., Xing, H. (2022). A novel multi-attention fusion network with dilated convolution and label smoothing for remote sensing image retrieval. International Journal of Remote Sensing, 43(4): 1306-1322. https://doi.org/10.1080/01431161.2022.2035465