J Inf Process Syst, Vol.21, No.3, pp.284~295, June 2025
https://doi.org/10.3745/JIPS.02.0224

ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

The Application of DCRN Algorithm Based on Attention Mechanism for Classification in Media Images

Peipei Dai*

Abstract

With the development of the information society, the significant increase in images in online environments poses challenges to media image management. To adapt to the development trend of the big data era and improve the classification effect of media images, this study introduces a dense connection refinement network (DRCN) in convolutional neural network image recognition, combined with attention mechanism, to fully utilize image features of different scales and improve the judgment accuracy of object detection based on increasing feature reuse times. The results indicate the consistent loss value of 0.08 for the DRCN-Attention model, while achieving a peak recall rate of 85% after 30 iterations, and a mean average precision exceeding 80%. The classification accuracy of ships reaches 82%, which is 8% higher than the support vector machine model. This indicates that the proposed media image classification method has high classification accuracy, provides a new technical reference for the field of computer vision, and has certain application value in the intelligent management of media images in the era of big data.

Keywords

Attention Mechanism, Convolutional Neural Network, DCRN, Media Images, Multi-Label Classification

1. Introduction

The key areas of image classification include real-world recognition, medical diagnostic recognition, and scene recognition, all of which are achieved by associating various labels [1]. Among many deep learning-based image classification methods, convolutional neural networks (CNNs) are considered as the most effective means due to their powerful feature extraction and mapping capabilities. CNNs can achieve accurate image classification by preserving all essential image data while decreasing computational complexity. However, CNNs encounter challenges, including slow inference speed and difficulty in fine-grained classification [2,3]. In addition, with the rapid development of the Internet, the amount of social information and the number of media images are increasing. These media images are not only huge in number, but also complex in nature, and irrelevant regions in the neural network media image classification model is introduced, which adopts an attention mechanism and a dense relational connectivity refinement network [4,5]. The main achievement of this research is to improve the effectiveness of image classification, which provides a solid foundation for the efficient management of media images in the information age [6].

Manuscript received July 11, 2023; first revision September 22, 2023; accepted November 6, 2023.

* **Corresponding Author:** Peipei Dai (daipeipei@wxit.edu.cn)
Department of Fundamental Courses, Wuxi Institute of Technology, Wuxi, China (daipeipei@wxit.edu.cn)

This work aims to convert human visual understanding into image data, which will allow computers to recognize images more efficiently. By mimicking the human visual system, computers will be able to process a large amount of visual resources. This breakthrough provides a new method for managing the growing and complex world of media images. It allows for improved image classification in the future to meet the growing demand for digital landscapes. This method can process large volumes of media images, including complex image types. Whether the image has a complex background or needs fine-grained recognition, the model can effectively classify. In addition, this approach allows computers to more effectively mimic the human visual system and process more image resources. This means that computers can use this approach to understand and classify images regardless of their content, background or complexity. In summary, this analysis offers a novel and efficient technique for image classification, as well as a potent tool for future image processing and analysis undertakings.

2. Related Work

Attention mechanism improves accuracy by assigning corresponding weight size based on the importance of information, providing new ideas for the development of artificial intelligence. Budak et al. [7] presented a fresh screening technique for the coronavirus. A SegNet network centered on the attention mechanism was proposed to enhance the precision of coronavirus computed tomography (CT) detection. The experimental results showed that the sensitivity of this method was as high as 92.73%. Huang et al. [8] focused on pedestrian attribute recognition and proposed a pedestrian attribute recognition network incorporating temporal attention mechanism based on image sequences. The empirical results demonstrated the effectiveness of this approach in various performance indexes. Sangeroki and Cenggoro [9] conducted a study that aimed to detect chest diseases in X-ray images through the attention mechanism with a light-tuple convolutional network. Results from experiments demonstrated that the proposed model significantly enhanced the detection efficiency and overall model performance. Zhang et al. [10] elaborated the goal of facial attribute editing. A multi-attention-based generative adversarial network was proposed. Experiments demonstrated that this method had the ability to improve correlations among attributes and precisely balance attribute editing while still preserving detailed features. Alhnaity et al. [11] proposed a recurrent neural network based on long short-term memory and attention mechanism to address the high error rate of multi-step prediction. Through experimentation, this method exhibited superiority in reducing error levels.

Media images are one of the important carriers of information dissemination in the Internet era, and their utilization is a hot topic in various fields of society. Based on the surge in demand for media image retrieval in the digital age, Olaode and Naghdy [12] proposed an automatic image annotation model that used machine learning for semantic indexing of media images. The model showed that updating the vocabulary helped to improve the system performance. Kumar and Ganesh [13] aimed to explore the security of media images and proposed an image encryption technique with RSA (Rivest-Shamir-Adleman) algorithm to encrypt. The results suggested that this method could enhance computational efficiency and reduce media image data leakage, making it useful for secure use of media images on open networks. Mao et al. [14] summarize the weak target localization of media images in the Internet of Things (IoT) environment and proposed a localization method based on multi-information fusion. When applied to a model with metallurgy as the target object, this method demonstrated strong feature resolution with superior performance advantages. Mo and Sun [15] designed an artificial intelligence image detection system based on IoT to address the slow speed and shallow image analysis of traditional media image detection methods. According to experiments, the system accurately detected media images, improved detection efficiency and system stability, and had high application value. Xin et al. [16] analyzed the effects of the widespread use of pornographic media on children in the Internet age. A region-based method for recognizing media images was proposed. The experimental results showed that the image recognition accuracy of this method reached 97.52%, which proved the effectiveness of the method. Patel et al. [17] aimed to study human motion recognition in the field of computer vision and designed six fusion models that incorporated the concept of feature fusion. Through experimental comparison, it is found that all feature fusion models significantly optimize the performance of human motion recognition models. Bhatt et al. [18] introduced the extensive application of computer vision in the information age. The significance of CNNs in the image processing field was emphasized, and the research direction and application scope of deep CNN in the future were discussed.

From the above results, it is clear that attention mechanisms often play an important role as optimization methods in the field of visual information processing due to their advantageous nature. The Internet era has introduced new themes for media images, especially new challenges in the development of image recognition and detection technologies. Therefore, the study proposes a CNN media image classification model based on attention mechanism and deep convolutional recursive network to solve the problems caused by huge image data, optimize the media image classification effect, and provide new ideas for media images to be better utilized in the Internet environment.

3. DCRN Media Image Classification Method based on Attention Mechanism

3.1 Hierarchical Features of CNNs Incorporating Attention Mechanism

CNN is a feed-forward neural network, which consists of multiple convolutional, pooling, excitation and fully connected layers. It extracts image features and enhances network expression through convolution and pooling operations, efficiently classifying media images. By incorporating channel attention into the sequence and excitation networks (SENet) of the CNN, the importance of feature channels following the convolutional operation is autonomously measured through continuous network training iterations. The weights are then ranked based on this measurement, and feature channels with high importance are given higher weights. This approach highlights the main features of the image. In the CNN, the input is set as y, the number of feature channels is C', the height of the feature map is H', and the width of the feature map is W'. Then, the dimension of the input feature map is $H' \times W' \times C'$. At a certain point in time t, the output U is obtained by convolution operation on the input Y. The height, width and number of features of the output feature map are H, W, and C, respectively, and its dimension is $H \times W \times C$. The global mean pooling operation is performed on the output U using the squeeze module in SENet, as demonstrated in Eq. (1):

$$Z_c = F_{sq}(U_c) = \frac{1}{W \times H} \sum_{m=1}^{W} \sum_{n=1}^{H} u_c(m,n), \tag{1}$$

where Z_c represents the computed result after image pooling under the attention mechanism. F_{sq} represents the squeezing process. (m,n) represents the coordinate positioning of each point in the feature image. $u_c(m,n)$ represents the feature information of each point. U_c represents the feature map of the intermediate result C.

After the squeezing operation, the input U dimension is changed to $1 \times 1 \times C$, the excitation module in SENet is used to learn the weight features and extract them, and the intermediate parameter is set to r. The parameters of the whole excitation process are adjusted. In the two-layer fully connected network, the number of channels C is compressed to C/r, and the number of channels is restored to the original data after reducing the number of parameters, i.e., C. The result after excitation is obtained by operation, which is calculated as shown in Eq. (2):

$$s = F_{ey}(z,W) = \sigma\big(g(z,W)\big) = \sigma\big(W_2 \mathrm{ReLU}(W_1 z)\big), \tag{2}$$

where F_{ey} represents the excitation process. z represents the squeezing result. σ and g represent the sigmiod function and the fully connected operation, respectively. W_1 and W_2 represent the weight matrix after the fully connected operation in two layers. ReLU represents the activation function. Assuming that the number of feature extraction layers is k, y_k is the result of the feature extraction layer in the k layer. The result y_{k+1} of the neural network in the $k+1$ layer is calculated, as shown in Eq. (3):

$$y_{k+1} = h(y_k) + F(y_k, W_k), \tag{3}$$

where $h(y_k)$ denotes the residual part of the neural network. W_k denotes the weight matrix of the current layer. $F(y_k, W_k)$ denotes the operation on the y_k layer. The feature extraction results for each layer are compiled into a feature map matrix. The splicing operation uses each component prediction to create an integrated prediction of the final result. The expression is shown in Eq. (4):

$$\mathrm{result} = concat(y_1, y_m, \dots, y_i), \tag{4}$$

where $concat$ denotes the matrix stitching operation that expands and merges the feature extraction results of each layer based on the row dimension. The feature maps of different layers are operated through the squeezing function F_{sq} to obtain the layer feature attention results, which are calculated, as shown in Eq. (5):

$$Z_c = F_{sq}(R_c) = \frac{1}{W \times H} \sum_{m=1}^{W} \sum_{n=1}^{H} r_c(m,n), \tag{5}$$

where $r_c(m,n)$ denotes the feature information in the current feature image. R_c denotes the feature map with the number of different layers c. Based on the stimulation operation, the output results are placed in two convolutional networks for a 1×1 convolutional network operation $conv$, which learns and extracts the weights. The activation sigmoid function is utilized to acquire the distinct weights of various layers. Different weights are assigned to each layer feature based on the original layer feature maps through certain operations, as shown in Eq. (6):

$$s = F_{e\psi}(z,W) = \sigma\big(conv(z,W)\big) = \sigma\big(W_2 \mathrm{ReLU}(W_1 z)\big), \tag{6}$$

where W_1 and W_2 denote the weight matrix after the 1×1 convolution operation. The stimulated results are input to the classifier. The final image classification result is obtained by fusing features between two fully connected layers in the classifier, as shown in Eq. (7):

$$r = \sigma\big(W_2 \mathrm{ReLU}(W_1 s)\big), \tag{7}$$

where W_1 and W_2 denote the parameter matrices of the fully connected layer.

3.2 DRCN-Attention Media Image Classification Model based on DCRN Algorithm Optimization

The deep convolutional recursive network, as a densely connected refinement network (DRCN), changes the connectivity between feature maps by adding densely connected blocks to enhance the correlation between each feature. The DCRN algorithm generates bounding boxes based on the features at different scales. This involves the dependence of features at each scale on the output of all previous feature layers, resulting in an extremely strong correlation. The number of connections of the densely connected blocks in the network A is calculated, as shown in Eq. (8):

$$A = \frac{X \times (X+1)}{2}, \tag{8}$$

where X denotes the total number of layers of the network. The target relationship module is constructed in the CNN to effectively learn the image features such as bounding box and appearance features to establish a relationship network between different image targets. This network is employed to assign weight to each target's category and location. The input target is N. m and n belong to the input targets. The regular image feature is f_B and the position feature of the image is f_G. Since different features have diverse locations, to avoid judgment errors caused by too divergent location values, the scale of the image needs to be unified by scale normalization and log operation, and the operation is shown in Eq. (9):

$$f_G' = \left(\log\left(\frac{|x_m - x_n|}{w_m}\right), \log\left(\frac{|y_m - y_n|}{h_m}\right), \log\left(\frac{w_n}{w_m}\right), \log\left(\frac{h_n}{h_m}\right)\right)^T, \tag{9}$$

where x and y denote the length and width of the feature, respectively. Based on the position features after coordinate transformation, the position feature weights are restricted by the $\max$ operation. Some cosine function and sine function operations are performed, and the weights of the m object on the position features are obtained for the current n object through the fully connected layer. The calculation is shown in Eq. (10):

$$w_G^{mn} = \max\{0. W_G, \varepsilon_G(f_G^m, f_G^m)\}, \tag{10}$$

where w_G^{mn} represents the position feature weight. The role of ε_G is to transform the coordinate information from four dimensions to high dimensions. For the image feature weight, the dot product operation dot, which is implemented through the fully connected layer, is used, and its operation is shown in Eq. (11):

$$w_B^{mn} = \frac{dot\left(W_k f_B^m, W_Q f_B^m\right)}{\sqrt{d_k}}, \tag{11}$$

where both W_k and W_Q are fully connected layer parameters that vary the dimensionality. d_k denotes the dimensionality after dot product. The position feature weights and image feature weights constitute the description of the total target features, and the relational weight of different total target features are calculated from w_G^{mn} and w_B^{mn}, as shown in Eq. (12):

$$w^{mn} = \frac{w_G^{mn} \cdot \exp(w_B^{mn})}{\sum_k w_G^{k_G} \cdot \exp\left(w_B^{k_B}\right)}, \tag{12}$$

where exp is an exponential function. The relational weight w^{mn} is a softmax operation combined with a 1×1 convolution, i.e., a linear transformation operation W_V on the input set $\{(f_B^n, f_G^n)\}_{n=1}^{N}$ to obtain the relational characteristics between the n-th target and the total target, whose expression is shown in Eq. (13):

$$f_R(n) = \sum_m w^{mn} \cdot (W_V \cdot f_B^m), \tag{13}$$

where f_B^m denotes the appearance features such as size, shape and color of the m-th input target, i.e., conventional image features. The operation produces relational features for each input target. These features are combined with the conventional image features and passed to the next layer of the network to complete the attention process, as calculated in Eq. (14):

$$f_B^n = f_B^n + Concat\left[f_R^1(n), \dots, f_R^{N_r}(n)\right]. \tag{14}$$

The primary function of $Concat$ in Eq. (14) is to merge the relational characteristics of numerous objectives. The structure of the target relationship module is shown in Fig. 1.

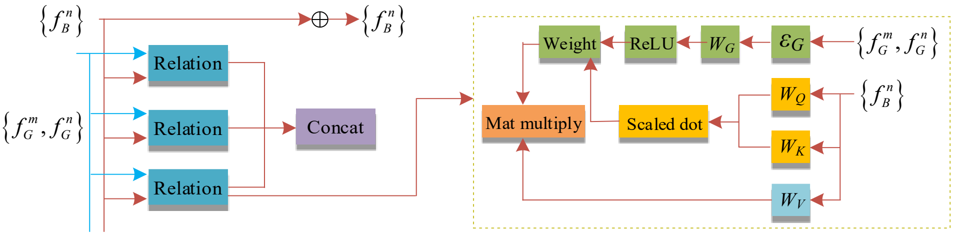


Fig. 1. Target relationship module structure.

The diagram in Fig. 1 demonstrates that each relationship module has two input features and multiple channels connected at each layer to learn various feature categories. The channels are utilized to extract target information with features in the target relationship module and consolidate all the target information for enhanced target detection accuracy. To avoid local maximum values misleading the final results in DCRN detection, the target relational network module is combined with non-maximum suppression to improve the network's adaptive ability to parameters. The input features are ranked based on the prediction scores of the frame classes, and the ranking information is fused with the image features to obtain the appearance features by up-dimensioning. The coordinate information of the predicted frames is used as input together with the appearance features. The relational features are obtained after target detection, which together with the feature scores constitute the final class scores. The structure of the DRCN-Attention media image classification model is shown in Fig. 2.

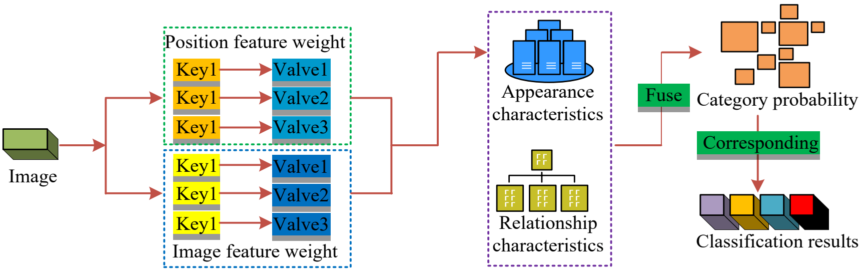


Fig. 2. DRCN-Attention media image classification model structure.

In Fig. 2, the DRCN-Attention media image classification model mainly consists of four modules: assigning weights, feature fusion, target detection and semantic correspondence. After acquiring the target image, image features and location features are extracted, and weights are assigned to them using the attention mechanism. These two are fused to get the relational features. The category probability is determined by relational features and original image features in target detection. Through semantic relational correspondence, different features are classified into corresponding categories to get the final multi-category labeling classification results.

4. Performance and Result Analysis of DRCN-Attention Media Image Classification Model

4.1 Hierarchical Features of CNNs Incorporating Attention Mechanism

The experiment was conducted on a server configured with NVIDIA Tesla P100 GPU. This server is equipped with 64 GB of memory and 1 TB of SSD hard drive. In terms of operating system, the Ubuntu 16.04 LTS is adopted. Python 3.6 is taken as the programming language and TensorFlow 1.12 is used as the deep learning framework for the experiments.

The experiment used two datasets, named the dataset A and the dataset B, respectively. The dataset A contains a large number of media images from different sources, including news, social media, and public image libraries on the internet. These images vary in size, color, and quality. This dataset is used to train and test the model to evaluate its performance in processing large-scale and diverse media images. The dataset B contains some images that require fine-grained classification. These images mainly include various types of plants and animals, as well as some common daily items. This dataset is used to evaluate the performance of the proposed model in processing images that require fine classification.

Mxnet is used as the framework, the initial learning rate is set to 0.01, and the weight decay is set to 0.0005. In two datasets, 80% of datasets A and B are randomly selected as the training set, 10% as the testing set, and 10% as the validation set. The experiment includes residual network (ResNet), support vector machine (SVM) and k-nearest neighbors (KNN) models for experimental comparison. Loss, mean average precision (mAP), and recall are selected as the testing criteria for the model. The loss value curves of different models are shown in Fig. 3.

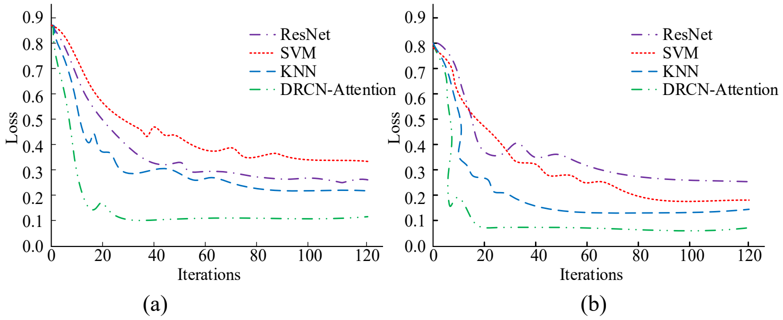
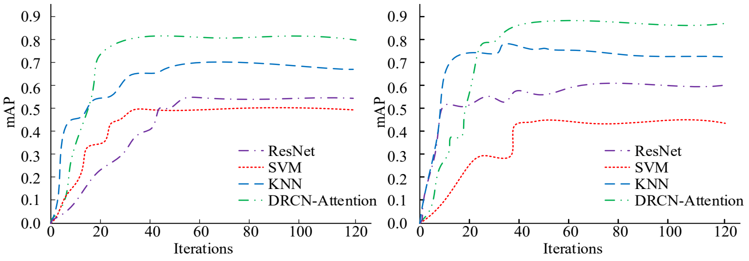


Fig. 3. Loss value curve of different algorithm models: (a) dataset A and (b) dataset B.

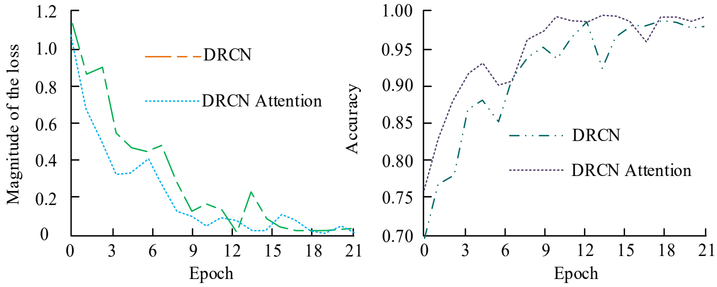
In Fig. 3, the DRCN-Attention model achieves a steady state loss value of 0.1 and 0.08 after 30 and 20 iterations for datasets A and B, respectively. The loss value fluctuates less, with only small fluctuations around 20 and 10 iterations. The ResNet model is better trained in dataset A than in dataset B, with 10 fewer iterations and a 0.05 reduction in loss value. The KNN model and the SVM model train better in dataset B, with lower loss levels but smaller variations. However, compared with the DRCN-Attention model in the unified dataset, the number of iterations increases by 10 and 50, and the loss values increases by 1.0 and 1.4, respectively. This indicates that the DRCN-Attention model converges faster and can reach a stable state with lower loss values in a shorter time, thereby improving the efficiency of the classification model. Fig. 4 displays the mAP of different models.

As depicted in Fig. 4, in dataset B, the DRCN-Attention model shows the highest mAP of 0.85 after the 40th iteration. The next highest mAP is the DRCN-Attention model trained in dataset A, reaching a mAP of 0.8 after 30 iterations. The mAP of both ResNet and KNN models produces better results, and their performance in dataset A is similar. The lowest mAP is the SVM model trained on the B dataset, which reaches 0.4, with a reduction of 0.45 compared to the DRCN-Attention model under the same conditions. The DRCN-Attention model is optimal on both mAP and variation. This indicates that the DRCN-Attention model is able to reduce the classification error, achieve high accuracy on multi-label classification, and maintain a stable state consistently, optimizing the model performance. The ablation experiment is performed on the DRCN-Attention model, that is, ablation experiments on the attention mechanism. It is compared with the DRCN algorithm. As shown in Fig. 5, the results demonstrate the effectiveness of the model.



(a) (b)

Fig. 4. mAP of different models: (a) dataset A and (b) dataset B.



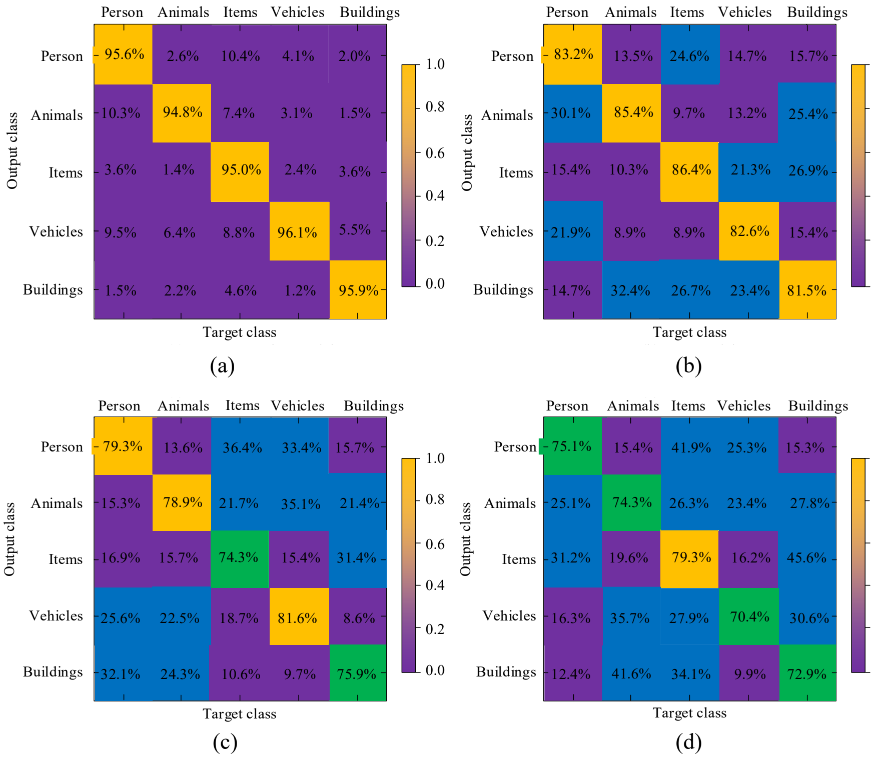
(a) (b)

Fig. 5. The ablation experimental results of the DRCN-Attention model: (a) loss and (b) accuracy.

In Fig. 5(a), the loss value of the DRCN-Attention algorithm is smaller than that of the DRCN algorithm during the iteration process, with an average difference of 0.1. As the iteration process progresses, the loss value of the DRCN-Attention algorithm decreases faster, eventually converging in the 10th iteration. In Fig. 5(b), the DRCN-Attention algorithm has the highest accuracy of nearly 99%, and the entire iteration process is superior to the DRCN algorithm, with an average improvement of approximately 0.5% and superior convergence.

4.2 Analysis of Classification Results of DRCN-Attention Media Image Model

The 5,000 high-resolution annotated images from PASCAL VOC dataset are selected for the practical application of the model, including five major categories of persons, animals, objects, vehicles, and buildings, and specifically 12 subcategories including birds, cows, bicycles, airplanes, ships, sofas, TVs, roads, railings, dams, men, and women. ResNet, SVM and KNN are still used as experimental comparisons. The classification accuracy of each major and subcategory is used as the test for model performance. Fig. 6 displays the confusion matrix of different models.



(a)

(b)

(c)

(d)

Fig. 6. Classification confusion matrix of different models: (a) DRCN-Attention, (b) KNN, (c) ResNet, and (d) SVM.

In Fig. 6, the classification accuracy of the DRCN-Attention model remains at 94.8% to 96.1%, with the highest classification accuracy. The objects with the lowest and highest classification accuracy are animals and transportation, respectively. The KNN model has the highest classification accuracy of 86.4% for items and the lowest classification accuracy of 81.5% for buildings. The SVM model exhibits the highest overall classification error, which remains within at 9.9% to 45.6%. Moreover, it displays the highest error probability in classifying buildings as items, at 45.6%. The best classification of items is 79.3%, with a decline of 16.8%. The fine class accuracy of different models is shown in Fig. 7.

As evidenced by Fig. 7, the DRCN-Attention model exhibits superior overall classification performance, with accuracy consistently maintained within the [0.75–0.82]. The transportation class yields the best and worst results for ships and airplanes, respectively, with classification performance displaying greater stability across all classes. The highest and lowest accuracy of the KNN model and SVM model are 0.62 and 0.74, respectively, where the KNN model has the best classification effect on TV and a larger recognition error on aircraft. Similarly, the best and worst objects classified by the SVM model are sofas and birds, respectively. It is evident that the DRCN-Attention model enhances the sensitivity and interest in target features, accurately detecting and classifying each category of objects in multiple labels, and improving classification accuracy.

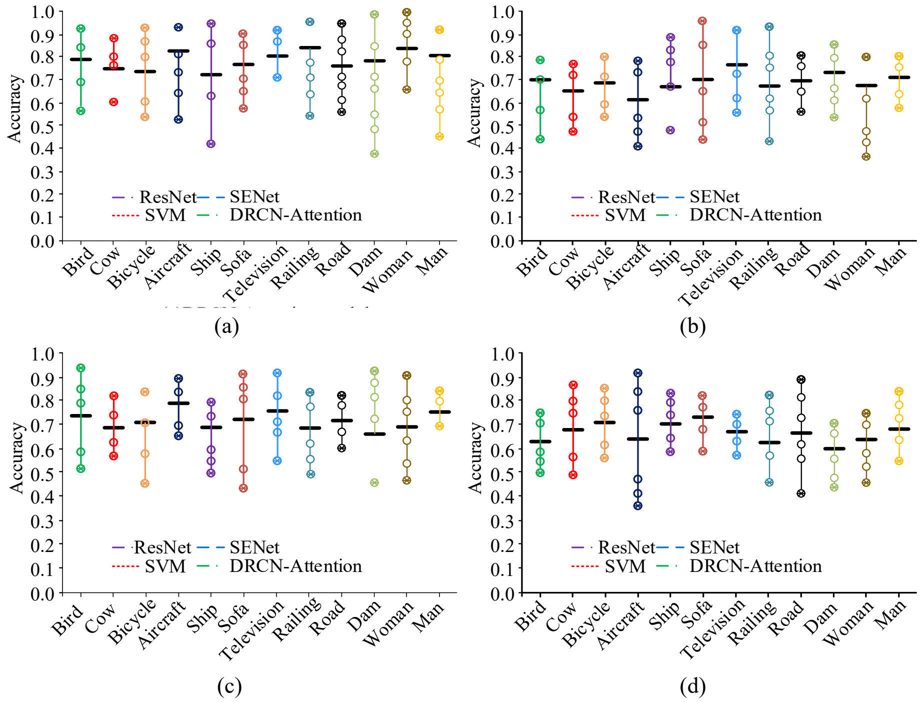


Fig. 7. Classification confusion matrix of different models: (a) DRCN-Attention, (b) KNN, (c) ResNet, and (d) SVM.

5. Conclusion

The CNN represents a major image processing technique in deep learning and plays a critical role in the field of multi label image classification. To optimize the classification effect of media images, this study proposed a CNN media image classification model combining attention mechanism and DRCN algorithm. The accuracy of the model construction index was verified. In application scenarios, the accuracy of the wide category is most suitable for the transportation category in the 94.8%–96.1%, and the accuracy of the subcategory remains in the [0.75–0.82], which is improved compared with the ResNet model of [0.04–0.1]. Additionally, ship classification exhibits the highest accuracy. Therefore, the CNN media image classification model based on the attention mechanism and DCRN algorithm improves the image classification accuracy, reduces the running procedures, accurately classify media images in a shorter time, and lays a foundation for the management and utilization of media images in the network. However, it has its limitations. First, although this model exhibits satisfactory results, it is limited to media image classification itself. Although there are different excellent algorithms to choose from, the main research focus is on media image classification algorithms. This obviously indicates that the application of deep learning algorithms in image classification needs further research and exploration.

Second, while the algorithm presented in this paper is capable of classifying and managing media images, it does not delve deeper into image retrieval. Therefore, for future research directions, potential extensions may focus on using image retrieval techniques to achieve more effective media image management. Based on the above considerations, other researchers can further explore how other deep learning algorithms can be applied to image classification, conduct more in-depth research on image retrieval techniques, and strive to achieve more efficient media image management, thereby expanding this research.

In this paper, a new partition-based device discovery scheme in lighting control networks is proposed. In the proposed scheme, all devices are divided into several partitions. In addition, to avoid collisions occurring due to multiple responses, each device sends a response message based on a response timer configured by the controller.

From the numerical analysis, the proposed scheme can provide a much lower device discovery time than existing schemes. Moreover, as the number of devices increases, the performance gap between existing and proposed solutions becomes increasingly large.

Conflict of Interest

The author declares that they have no competing interests.

Funding

None.

References

[1] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, “Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image,” *IEEE Transactions on Cybernetics*, vol. 49, no. 7, pp. 2406-2419, 2019. https://doi.org/10.1109/TCYB.2018.2810806

[2] F. Rustam, A. A. Reshi, W. Aljedaani, A. Alhossan, A. Ishaq, S. Shafi, et al., “Vector mosquito image classification using novel RIFS feature selection and machine learning models for disease epidemiology,” *Saudi Journal of Biological Sciences*, vol. 29, no. 1, pp. 583-594, 2022. https://doi.org/10.1016/j.sjbs.2021.09.021

[3] C. Li, C. Liu, L. Duan, P. Gao, and K. Zheng, “Reconstruction regularized deep metric learning for multi-label image classification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 7, pp. 2294-2303, 2020. https://doi.org/10.1109/TNNLS.2019.2924023

[4] A. G. Priya Henry and A. Jude, “Convolutional neural-network-based classification of retinal images with different combinations of filtering techniques,” *Open Computer Science*, vol. 11, no. 1, pp. 480-490, 2021. https://doi.org/10.1515/comp-2020-0177

[5] C. Wang and Y. Wang, “Opportunities for digital culture and innovation under cultural construction in the big data era: take digital cultural innovation in Jiangxi universities as an example,” *International Journal of Social Science and Education Research*, vol. 2, no. 11, pp. 38-43, 2020. https://doi.org/10.6918/IJOSSER.202001_2(11).0007

[6] S. K. Thangarajan and A. Chokkalingam, “Integration of optimized neural network and convolutional neural network for automated brain tumor detection,” *Sensor Review*, vol. 41, no. 1, pp. 16-34, 2021. https://doi.org/10.1108/SR-02-2020-0039

[7] U. Budak, M. Cibuk, Z. Comert, and A. Sengur, “Efficient COVID-19 segmentation from CT slices exploiting semantic segmentation with integrated attention mechanism,” *Journal of Digital Imaging*, vol. 34, no. 2, pp. 263-272, 2021. https://doi.org/10.1007/s10278-021-00434-5

[8] C. Huang, J. Pei, and Y. Zhao, “Pedestrian sequence attribute recognition method with multi-feature fusion combined with temporal attention mechanism,” *Journal of Signal Processing*, vol. 38, no. 1, pp. 64-73, 2022. https://dx.doi.org/10.16798/j.issn.1003-0530.2022.01.008

[9] B. A. Sangeroki and T. W. Cenggoro, “A fast and accurate model of thoracic disease detection by integrating attention mechanism to a lightweight convolutional neural network,” *Procedia Computer Science*, vol. 179, pp. 112-118, 2021. https://doi.org/10.1016/j.procs.2020.12.015

[10] K. Zhang, Y. Su, X. Guo, L. Qi, and Z. Zhao, “MU-GAN: facial attribute editing based on multi-attention mechanism,” *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 9, pp. 1614-1626, 2021. https://doi.org/10.1109/JAS.2020.1003390

[11] B. Alhnaity, S. Kollias, G. Leontidis, S. Jiang, B. Schamp, and S. Pearson, “An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth,” *Information Sciences*, vol. 560, pp. 35-50, 2021. https://doi.org/10.1016/j.ins.2021.01.037

[12] A. Olaode and G. Naghdy, “Review of the application of machine learning to the automatic semantic annotation of images,” *IET Image Processing*, vol. 13, no. 8, pp. 1232-1245, 2019. https://doi.org/10.1049/iet-ipr.2018.6153

[13] S. Kumar and D. Ganesh, "Image cryptography using RSA algorithm," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 5, no. 4, pp. 835-837, 2021. https://journals.indexcopernicus.com/api/file/viewByFileId/1463843

[14] K. Mao, G. Srivastava, R. M. Parizi, and M. S. Khan, “Multi-source fusion for weak target images in the Industrial Internet of Things,” *Computer Communications*, vol. 173, pp. 150-159, 2021. https://doi.org/10.1016/j.comcom.2021.04.002

[15] C. Mo and W. Sun, “Point-by-point feature extraction of artificial intelligence images based on the Internet of Things,” *Computer Communications*, vol. 159, pp. 1-8, 2020. https://doi.org/10.1016/j.comcom.2020.05.015

[16] X. Jin, Y. Wang, and X. Tan, “Pornographic image recognition via weighted multiple instance learning,” *IEEE Transactions on Cybernetics*, vol. 49, no. 12, pp. 4412-4420, 2019. https://doi.org/10.1109/TCYB.2018.2864870

[17] C. I. Patel, S. Garg, T. Zaveri, A. Banerjee, and R. Patel, “Human action recognition using fusion of features for unconstrained video sequences,” *Computers & Electrical Engineering*, vol. 70, pp. 284-301, 2018. https://doi.org/10.1016/j.compeleceng.2016.06.004

[18] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, “CNN variants for computer vision: history, architecture, application, challenges and future scope,” *Electronics*, vol. 10, no. 20, article no. 2470, 2021. https://doi.org/10.3390/electronics10202470

Peipei Dai **https://orcid.org/0009-0002-5479-9965**

She received M.S. degree and Ph.D. in applied mathematics in Soochow University in 2011 and 2015, respectively. Since July 2015, she is a lecturer at the Department of Fundamental Courses of Wuxi Institute of Technology. Her current research interests include combinatorial design, combinatorial coding and mathematical modeling.