Smart Album Management System Based on SE-ResNeXt

Zhendong Feng, Wei Liu and Yinghuai Yu*

School of Mathematics and Computer Science, Guangdong Ocean University, China
E-mail: YingHuai_Yu@126.com
*Corresponding Author

Abstract

With the rapid popularization and development of smart phones and other technological devices, pictures have become the main media for people to record information. However, the traditional mobile photo album has many problems. First of all, with the development of the times, the higher the pixel of the image, the larger the memory required. Obviously, the traditional file storage structure can no longer meet the storage of users' massive photos. Secondly, people store a large number of face images in mobile phones, so there is a strong demand for face recognition and classification management based on different faces. Third, in the face of the management of massive photos, general image recognition and classification is also a very demanding function. In response to the call of "deeply implementing the digital economy strategy" in today's era, our team makes full use of the functions of the cloud platform and a large number of industrial resources, and integrates independent optimization algorithms to develop an intelligent cloud album management system that realizes intellectualization and application innovation. SE-ResNeXt algorithm is the core algorithm of this

system, which can recognize and extract effective information from massive images in various application scenarios, and help users to intelligently and automatically classify and manage images according to different contents. This paper deeply studies the Intelligent Cloud album management system based on SE-ResNeXt. The system is built by nginx+uwsgi+django+vue as a whole. It has the functions of intelligent classification, face recognition, cloud storage and so on. It aims to provide users with simpler, more intimate and more intelligent album management services.

1 Introduction

With the continuous improvement of smart technology devices in the camera function and the high pixel camera is generally standard, "camera" gradually into thousands of households, humans are gradually getting used to the camera to record daily life, whether it is a glimpse of the journey alone, or in the most ordinary moments of life, should be remembered by the image. However, due to the increasing number of images generated or recorded by users today, directly resulting in a huge consumption of space on the device.

According to data analysis, the number of photos taken by users worldwide has exploded exponentially. After photos are generated, more than three-quarters of Asian users save them directly in their mobile albums, creating a significant memory challenge for mobile devices. The report shows that about 8.5% of Asian users store photos in the cloud, and 13% of users in China upload to the cloud [1]. Cloud Storage can provide enough memory space, prevent data loss, and protect photos in a more cost-efficient way, which will undoubtedly become a major trend in future storage methods.

In addition to this, the user's demand for higher and higher photo quality, cell phones are generally configured with high-definition cameras, not only to bring users a better experience, but also to bring new opportunities for the picture application class system, with the picture quality has been the development of the picture application system industry, today's cloud photo album only storage function has long been far from enough. With the promotion and popularity of cloud computing technology, Cloud Storage has gradually become an effective way to solve the problem of enterprise mass data storage and management [2]. The cloud photo album application system market has unlimited possibilities and is showing a spurt and powerful vitality.

Through the research on today's photo album management market, the photo album management function in smart phones is relatively simple, and collecting photos occupies a large amount of memory, which affects the performance of mobile phones. However, if the pictures are transferred to the cloud disk storage software similar to Baidu online disk, many operations on photos are missing. Therefore, various cloud intelligent album management systems are emerging. Among the intelligent album management systems with cloud services, Google album, Xiaomi cloud album and oppo cloud album are better known. They all provide Cloud Storage services. In Google album, It includes the basic operation of pictures and some interesting functions, such as collage photos, making photo albums, etc, while Xiaomi cloud photo album and oppo cloud photo album only provide batch operations such as photo display, recycle bin, upload and download.

In view of the product analysis of the current PC intelligent album management system and the shortcomings of various cloud album software or systems in the current market, the intelligent cloud album management system based on SE ResNeXt in this paper has been optimized for many times, and has developed cloud storage, face detection, general image recognition and classification, automatic video editing and other functions close to user needs [3]. The intelligent cloud album management system in this paper has been used in Windows10 system for competitive product data analysis with popular Xiaomi albums, cloud albums, Baidu albums and other products in China through different use scenarios. The data after repeated upload tests of nearly 20000 photos of different types using Apache Jmeter, the official testing tool of the Apache Foundation, The average response speed of this system is about twice that of other similar systems of the same brand, which effectively solves the practical pain points of the current album system, such as slow image processing speed and low degree of intelligence. The fast and accurate intelligent image processing technology greatly improves the efficiency of users using the album system.

2 Overall System Design

The system mainly includes photo module, general album module, smart album module and other functional modules, and its overall architecture is shown in Figure 1.

Users can upload multiple pictures locally, and then put them into the photo module for display. In addition to the basic operations of downloading, deleting and collecting, they can also customize general albums according
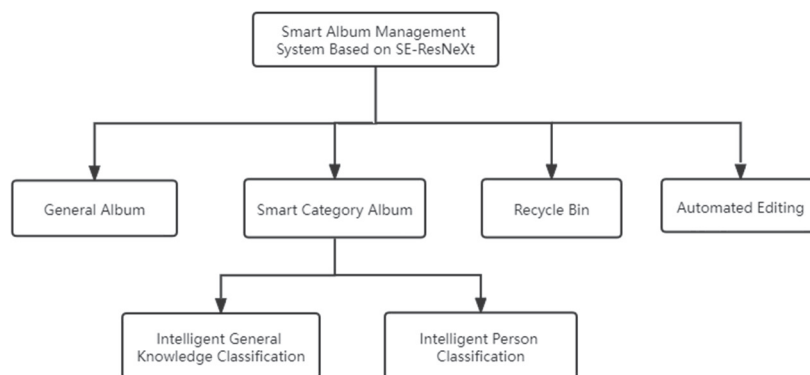
Figure 1   Overall system design.

to their preferences [4]. Intelligent photo album includes intelligent common sense photo album, intelligent character classification photo album and intelligent location classification photo album. . . After users successfully upload pictures, they can automatically extract picture features and classify common sense pictures (buildings, people, places, animals, etc.). The system also provides functional modules such as clustering classification, positioning classification, manual error correction and GIF generation.

3 Algorithm Design

In this system, face recognition related technology is used to detect whether there is face information in the photo, check the part and size of the face in the image if there is a face feature, then cut out the face screenshot, compare the face features, and gather the face pictures that reach a certain degree of similarity into a picture library. Before solving the current problem, the problems are the size of the face in the photo and the low resolution of the photo. Face occlusion by external objects, shooting angle, lighting, expressions, and other factors can pose a huge challenge to Face Detection.

Face Detection in layman's terms is to input a picture and then find out all the face positions in the picture, often using rectangular boxes to frame each of the recognized faces. Then a number of faces are output with position information similar to (x, y, w, h). The development process of Face Detection algorithms can be roughly divided into three stages: early algorithms, the AdaBoost framework, and the deep learning era [9].

The early algorithm stage is based on geometric features, subspace-based and many other types of algorithms. Algorithms based on geometric

analysis, human face are composed of five senses, because the shape, size, distance and many other differences of the five senses and make each face in the world have differences, so these influencing factors such as: the relationship between the features and structure of the face is an important feature of face recognition. The relationship between these features is used to compare the unknown face feature vector and the detected face feature vector to determine the degree of matching. The subspace algorithm, for Face Detection technology, treats all pixels containing faces as vectors of dimensional subtraction, and then obtains low-dimensional vectors by projecting the dimensional subtraction vectors in a low-dimensional space to achieve a better discrimination between different faces. The second stage is to compare and distinguish different faces by machine learning algorithms according to the idea of artificial features + classifier, through the training of a large number of samples to build a pedestrian Face Detection classifier, the main extracted features are edge, texture, gray scale, gradient histogram map and other information, the classifier mainly contains neural SVM, neural network, adaboost [6].

The face recognition process is shown in Figure 2. Firstly, it detects whether the image contains faces. If so, it extracts the unstructured face information of the people contained in the image file in the form of structured fields. Then the algorithm extracts the faces in the image in turn and stores them in the local area [7]. At the same time, it records which image they come from, and then compares them with each representative face in the specified cloud face database. We take out the faces stored in the local database and perform "similar face search" for each representative face in the specified cloud face database. If the returned list is obtained, it indicates that there are similar faces, and a representative face with a maximum confidence level of no less than 72% is obtained; If the returned list is not obtained, or there is no representative face with a confidence of no less than 72%, it means that the representative face does not exist in the representative face database, so a new face database will be created for the face. If the information from the previous step is obtained, which means that the face used for this comparison is from the same person as one of the representative faces in the face database, the face is stored in the corresponding face database to which the representative face belongs [8]. Since then, the recognition process of face recognition album has been completed.

Image classification is to input the image into the computer in order to determine the category of the image. Its core goal is to assign a label to the image from a set of determined labels, so as to realize the automatic
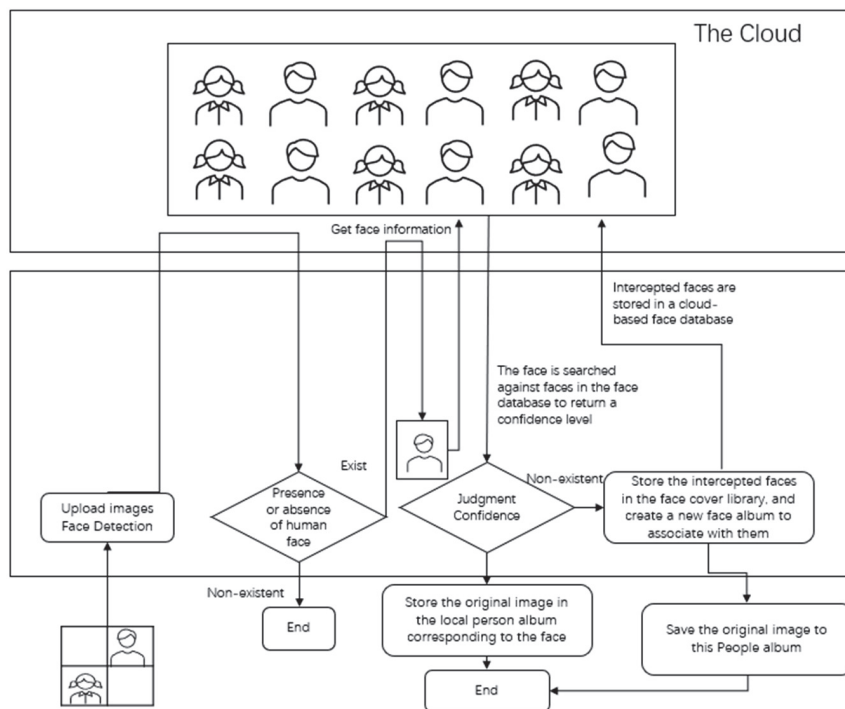
Figure 2   Face recognition algorithm process.

recognition and archiving of the image. When photos are input into the computer, the computer image is a large amount of data combination of a three-dimensional array, and the size of the array changes from wide to wide × high × 3, 3 represents red, green and blue channels. Generally speaking, the size of the image array obtained is very large, which contains a large number of array elements. The range of elements is 0–255, 0 is all black and 255 is white. Each array element is a digitally converted color. Image classification algorithms are widely used in reality, such as target recognition in traffic scenes and common sense image classification in intelligent photo albums. It can complete the automatic recognition and archiving of images, and can be classified from the types of photos, including people, animals, scenery, transportation and buildings [6,9].

In this paper, the image intelligent classification of the intelligent cloud album management system based on SE-ResNeXt adopts the semantic segmentation technology for quantitative analysis of different types of images, divides the image into a series of defined sub images by region, extracts the
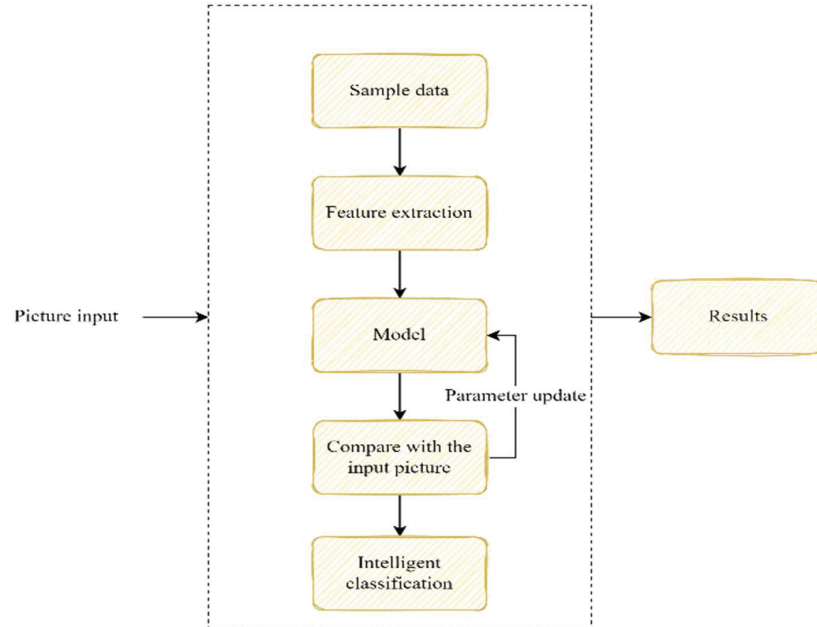
Figure 3 Image intelligent classification algorithm flow.

features, and matches the parameters with the original image, thus realizing the image intelligent classification of the cloud album management system. The algorithm flow is shown in Figure 3.

The traditional image classification process is shown in Figure 4: first, we input the processed data set and classify and mark the data set images. The test set contains various types of images that we need to segment after creating the data set. We need to preprocess the data. Preprocessing is mainly to improve the quality of the image and enable the computer to better train and learn the characteristics of the data. The main preprocessing operations are image size adjustment, image normalization and image denoising. Because there are a lot of redundancy and noise in the bottom features, which reduces the robustness of the feature representation [10]. In order to improve the robustness of the feature representation, the following features should encode the bottom features through the feature transformation algorithm, and perform the constraint operation after the feature coding is completed. The maximum or average value of each dimensional feature is obtained in the spatial range to understand the feature expression without feature distortion. The computer obtains the features of the image, then classifies
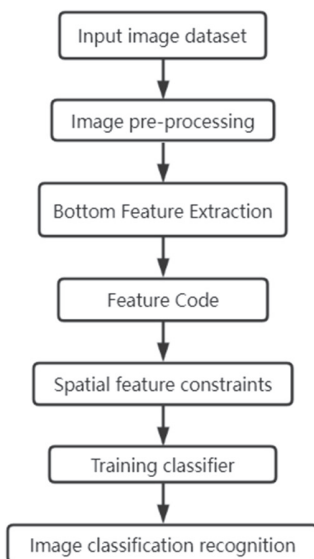
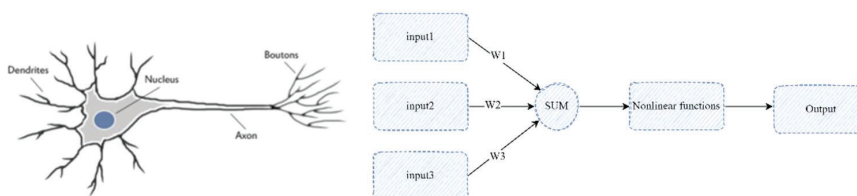**Figure 4**   Traditional image classification process.

Figure 5   Neural networks.

it into categories with similar features, and inputs the data features extracted from the data set into the classifier for training and learning.

The basic idea of deep learning is to learn the hierarchical representation of features, and describe the images in an unsupervised or supervised form [11]. The neural network shown in Figure 5 includes a large number of brain like neurons. The neuron will receive the input signals, then multiply them, add them with the corresponding weights, and input them into the nonlinear function. Neural network is similar to a good function, which can be used to understand everything of the system, but it is difficult to find such a function, so it is necessary to evaluate the neural network. With the rapid development of modern technology, more advanced computers, larger data sets and deeper technical training ability are no longer problems, and the popularization of deep learning has made a breakthrough.

Compared with other image classification algorithms, the image classification algorithm based on deep learning can obtain more abstract and advanced image information. The convolutional neural network often mentioned in deep learning can simplify the image preprocessing process. In particular, when manual image preprocessing is required, training can be started by directly inputting the original image. So far, convolutional neural network has been widely used in various image processing. Lenet-5 network model has been applied to practice for the first time. The academic interest of convolutional neural network began with the proposal of lenet5 network, but its scale is small and is not suitable for large-scale data training [12]. In recent years, the advantages of convolutional neural network in the field of image classification have become more and more prominent. Scholars have studied and improved the model from all aspects to continuously improve the performance of the model [13]. At present, there are many excellent convolutional neural network models, such as alexnet, VGg net, googlenet, RESNET, densenet and so on, which have been recognized and widely used in academic circles. The complete convolutional neural network is shown in Figure 6.

In the selection of neural network models, we studied and compared Alexnet, VGG16, Googlenet, Resnet and SE-ResNeXt models. Alexnet has two main features.

The sigmoid activation function is replaced by the relu activation function, which makes the computation much less and also converges faster. the mathematical expression of Relu is shown in Equation (1).

$$f(x) = \max(0, x) \tag{1}$$

To solve the problem of model overfitting, its Dropout layer is added after each fully connected layer. there are larger convolutional kernels in AlexNet (11×11, 5×5), the number of channels in the network is 64 at first, and then it grows exponentially after each pooling layer, and the size of the feature map decreases exponentially.
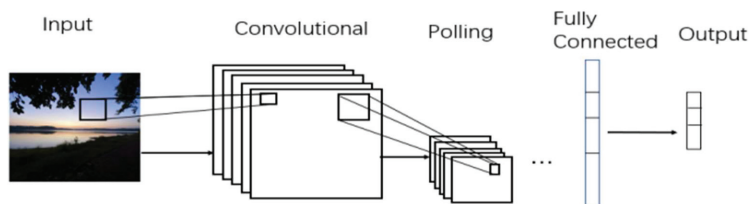

Figure 6   Convolutional neural networks.

Through the trained data analysis, VGG16 has a good effect on ImagNet. However, when using VGG16 for training, it takes a long time and has high requirements on device performance. In addition, there are a large number of channels in VGG16 convolution layer, so the model efficiency is not high. The obvious feature of Googlenet after the optimization of the original model is that it replaces the full connection layer with the global average pooling layer after the last convolution layer. Googlenet was about 3.3 percent more accurate on ImageNet, and training was more efficient overall than VGG16 [14]. Similar to the processing method of Googlenet, the SE ResNext model refers to the processing method of Googlenet in the training and tuning process to set the window size of avg pooling to the size of feature map, and uses the channel of feature map as a breakthrough to make the last layer convolution output channels with the same number of categories, and then pool them on a global average to output vectors with the same length as the number of categories, so as to avoid the risk of over fitting caused by full connection, In addition, it can achieve the same transformation function as full connection, thus improving the structure of the model and realizing the use of regularization to prevent over fitting in the multi round training process. Therefore, SE ResNext finally chose the pooling optimization method of the global average pool layer. The comprehensive use of residual module can train the 152-layer residual network, and its accuracy is higher than VGG16 and Googlenet, and its calculation efficiency is also higher than the former two. In this paper, we set the Crop Size to 224 and the Resize Short Size to 256 in the T4GPU based environment. We test the 152 layer SE ResNeXt network model with Batch Size = 1, Batch Size = 4, and Batch Size = 8 in FP16 and FP32 respectively. We also record the values of top1 and Top5 in the experiment in detail. From the experimental data record in Table 1, we can see that the Top5 value of 152 layer SE ResNeXt network model can reach more than 92.7%.

In this paper, according to the non maximum activated double branch network, it includes non maximum activated module and isomorphic double

Table 1 Top 1 and Top 5 values of 152 layer SE-ResNeXt network model under different batch size values of FP16 and FP32. The crop size is 224 and the resize short size

Index	Parameters	Batch Size = 1	Batch Size = 4	Batch Size = 8
Top1	FP16	67.54%	72.37%	76.21%
	FP32	83.67%	87.47%	89.24%
Top5	FP16	69.42%	75.98%	79.59%
	FP32	86.51%	94.15%	97.51%

branch sub network. The non maximum activated module is used to output the maximum activated feature and non maximum activated feature. The output feature is input into the isomorphic double branch sub network, which is learned by the isomorphic double branch sub network and outputs the image feature of the target to be classified [5, 8, 15]. When the number of cycles does not reach the preset number, the weight matrix, the category weight matrix and the Laplace matrix of the target domain are updated respectively according to the label and the pseudo label to obtain the updated weight matrix, the updated category weight matrix and the Laplace matrix of the updated target domain. Then, a very important concept is pseudo label – let the number of cycles reach the preset number, so as to obtain the classification result. The classification result here is the result of the classification of each sample image in the target domain. This paper selects five algorithm models, namely, Alexnet, Densenet, VGG16, Googlenet, and SE ResNeXt-152, to conduct experiments on setting the label weight matrix parameters of the observation results of the number of cycles under the same hardware device and network status. Based on the analysis and comparison of the experimental results, the label weight matrix parameters of SE ResNeXt-152 model are more in line with the expected requirements, and when the set value of the number of cycles is greater than 45, the speed advantage of the model in processing data is gradually highlighted, Finally, SE ResNeXt-152 model is selected as the basic algorithm model of this paper. In the SE-ResNeXt network model, the focus is on the channel relationship, and based on this, a "sequence and exception" (SE) block is generated. The channel direction is adaptively recalibrated by explicitly modeling the interdependence between channels [16]. Its core idea is to learn the similarity between the front and back frames, so as to complete the matching between the template frame and the search frame. Among them, similarity learning is a key link that affects the performance of tracking algorithm. Taking the similarity learning of twin networks as the starting point, the existing similarity learning method of deep cross-correlation (dw-xcrr) is improved, and a multi-scale similarity learning algorithm is proposed. The deep feature layer and shallow feature layer of the model are fused, and then the obtained features are fused with the deep feature layer. The channel attention mechanism is added to the two-way fusion to enhance the semantic information. Inputting the sample images in the pre acquired initial sample set and the new sample set into the initial image classification model to generate the first loss value corresponding to the initial sample set and the new sample set respectively, wherein, The initial image classification model includes a full connection layer corresponding

to the new sample set, a backbone network trained based on the initial sample set and a full connection layer corresponding to the initial sample set. The first loss value is generated based on the results output by the full connection layer corresponding to the initial sample set and the new sample set and the corresponding sample annotation information respectively, The weight associated with the generated first loss value in the initial image classification model is adjusted to train the initial image classification model. This embodiment improves the scalability of the training image classification model [18]. SE-ResNeXt operates on inputs by learning branch structures to better represent features. Structurally, it uses a branch to learn how to evaluate the correlation between channels, and then operates on the original feature mapping to calibrate the input. With the help of branch learning, it can get a more suitable neural network representation. The method proposed in this paper can extract the target region under different image backgrounds in a variety of environments, make full use of gray moment, color moment and invariant moment, extract the shape, color, color and other features, analyze the texture and other features of the image on the basis of obtaining the region of interest of the image, and has certain anti-noise ability, In essence, it is closer to the process of human visual attention.

SE-ResNeXt is in essence a compression and excitation network, Squeeze-and-Excitation: feature rescaling volume (or called attention mechanism). Specifically, it automatically acquires the importance of each feature channel by learning, further enhances useful features based on the acquired importance, and suppresses features that are less useful for the task at hand [18]. the SE module is shown in Figure 7.

Based on the SE-ResNeXt network model, the first step is to perform the Squeeze operation, which is the compression part of the SE-ResNeXt network model. In order to obtain the global perceptual field, it must be collected globally, and the size of the output image corresponds to the number of input resource channels, representing the global response distribution in
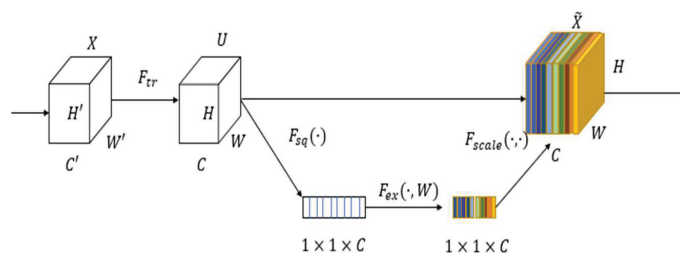

Figure 7   SE module.

the resource channels. The size of the original feature map is H*W*C, where H represents the height, W represents the width, and C represents the number of channels. In the compression operation, H*W needs to be compressed into one dimension [19]. In other words, H*W*C needs to be compressed into 1*1*C. This method is equivalent to obtaining an overall view of H*W before compression, making the sensory region wider. The obtained statistical value Zc is obtained by compressing the region with spatial dimension H X W. This is indicated by the following formula:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{w} u_c(i, j) \tag{2}$$

Next is the Excitation operation, the meaning of this operation is to allow each resource channel to establish a certain connection between them, the establishment of this correlation is essentially done by generating weights for each resource channel through the complete connection layer. Each resource channel outputs the weights after selecting the importance of the resource, but this is only the beginning, because after this is followed by a one-by-one weighting process of the previously obtained resources, the actual meaning is to complete the initial recalibration process of the resource-to-channel dimension [20]. After obtaining the 1*1*C representation of the squeeze, all that has to be done is to predict the importance of each channel. A common implementation is to add a fully connected layer to perform the prediction, and the next operation after the prediction is completed is to stimulate the corresponding channel of the previous resource graph, and then perform the subsequent operations. In order to better show this process, a threshold mechanism in the form of sigmoid is chosen for the study in this paper. The mechanism using the following equation.

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_z \delta(W_1 z)) \tag{3}$$

In the study of the optimization scheme, in order to reduce the complexity of the whole image classification model and improve the generalization ability, a bottleneck structure containing two Fully Connected Layers (FCL) is finally adopted through multi-program multi-dimensional research and analysis, where the first Fully Connected Layer plays the role of dimensionality reduction with a dimensionality reduction factor of r, and then ReLU activation is used [21]. The final Fully Connected Layer restores the original dimensionality, followed by the ablation experiment, and after obtaining the

gating unit, the output vector product.

$$\check{X}_c = F_{scale}(u_c, s_c) = s_c u_c \qquad (4)$$

Through the training and verification of SE-ResNeXt model, deep learning is no longer concerned with what filters can be used to perfectly filter out the better-performing features, input the original photo, without knowing in advance what type of filter to use, assuming that the first use of 10 color filters, the results obtained and the expected inconsistency, then modify the filter at this time and try again, if not, continue to modify, after several attempts until the desired results, then finally the desired combination of filters can be obtained.

In the dataset collection, 6000 human images, 6000 animal images, 6000 landscape images, 6000 vehicle images and 6000 building images were collected by web crawler technique and online image acquisition, and a general classification dataset with a total number of 30,000 photos was built. We divided the 30,000 images into three modules: training set, validation set and test set according to the ratio of 6:2:2. Then we trained the designed SE-ResNeXt-based neural network with different cycles using the training dataset, and tested it using the test set, and completed the training with multiple hyperparameter data under different cycles to validate the model by calculating the accuracy. The experimental results are shown in Table 2 for training the model by setting EPOCHS values of 5, 10, and 20, respectively, and it can be seen that the loss value of training is getting smaller as EPOCHS increases and no overfitting occurs. At around EPOHCHS = 10, the difference between the loss values of both is not significant, but the accuracy of the test at EPOCHS = 20 is 3.4% higher than the accuracy of the test at EPOCHS = 10. After the model training is completed, we use the trained model and model weights to build a web application to complete the prediction of the specified images and successfully return the data related to the prediction results.

Table 2 Experimental results

Results	EPOCHS = 5	EPOCHS = 10	EPOCHS = 20	EPOCHS = 20
		Periodicity		
Accuracy(%)	79.7	89.3	92.7	97.5
Minimum loss value of training set	1.15	0.445	0.203	0.139
Test set minimum loss values	0.968	0.407	0.247	0.127

According to the test records, it can be analyzed that the accuracy of the model is 79.8%, 89.3%, 92.7% and 97.5% respectively for the training sets with epochs of 5, 10, 20 and 40. The minimum loss of the training set and the test set decreases with the increase of the epoch. Combining the training and verification data of the model, the following model optimization techniques can be obtained to improve the accuracy.

- Regularly adjust the relationship between learning rate and batch size. Theoretically, a larger batch size indicates that the gradient value calculated based on each batch data is more consistent with the entire dataset. Understanding from the mathematical level is to make the variance smaller. Therefore, when a more accurate new direction is determined, the number of samples selected for training can be increased. Generally speaking, to ensure the rationality of the model, the batch size should be modified at the same multiple as the initial learning rate.
- The epoch value is set from small to large for training. First, use a smaller learning rate to train fewer epochs, because the parameters of the network are randomly initialized, if a larger learning rate is used at the beginning, the training process is prone to numerical instability, therefore, the training process can be stabilized and then gradually increase the learning rate (Learning rate) to continue training. In general, when the other network parameters remain unchanged, only the epoch changes, then when the loss value increases with the epoch, the maximum epoch value of the current model can be roughly determined, so that the next operation can be targeted. Of course, if the value of echo reaches the threshold value and continues to increase its value, the model will often be over fitted in the training process. At this time, attention should be paid to adjusting the model structure to achieve better training effect.
- Do not change bias parameters at will, especially do not perform weight attenuation operation. The main function of weight attenuation operation in network model is to adjust the influence of model complexity on loss function, that is, to reduce the over-fitting phenomenon of SE-ResNext model by constraining related parameters of network layer.

4 Conclusion

In this paper, an intelligent album management system based on SE-Resnext is proposed. Combined with the mainstream image detection algorithm, the

training and testing of the model are verified. With the remaining parameters in the network structure unchanged, it can be found that with the increase of the EPOCH in the training of SE-Resnext model within the threshold, the training effect is gradually improved. In the process of training, we deepened the network structure SE-ResNeXt model to some extent to explore better effects. Theoretically, the original layer can learn the same parameters as the original model. Of course, the premise of this result is that the newly added layer is the identity mapping. In this way, the optimized deep model structure can show better results than the original model structure. In the actual process of training and validation data show that the increase of network layer after more than a threshold, the training error tend not to send a rose, if the training set the amount of data that is not enough, by using the method of cross validation, to a large extent can help us in selecting the optimal parameters of noise reduction, in other words, Better results should be found in the space of the new model solution than in the subspace corresponding to the original model solution. The image classification effect of the smart album system based on SE-ResNeXt is remarkable, realizing the functions of automatic classification of input images and detection of different objects. In Face Detection, through the continuous upgrading and optimization of the Face Detection model, the final accuracy of face recognition in the album system can reach 93.6%. In the later stage, the group will continue to study the upgrading optimization scheme based on SE ResNeXt model. For example, when FP is 32, the group will continue to adjust the Batch Size value to test the Top1 and Top5 values of SE ResNeXt, and continuously improve the corresponding rate of the model. In addition, the group will study the relationship between different parameter settings of the model and effectively reduce the lossz value in the training process, so as to further improve the accuracy of image classification, and integrate multiple high-precision target detection models, Through the intelligent album management system, the intelligent classification and accurate detection of images can be realized in complex situations such as low resolution.

Acknowledgement

References

[1] Miao Zhuang, Zhao Xun, Wang Jiabao, Li Yang, Zhang Rui, Xu Bo, Wang Yapeng, Yang Li, Zhao Xinxin, Yang Yixin. A fine-grained image classification method and system based on two-branch network [P]. Chinese Patent:CN113902948A, 2022-01-07. https://doi.org/10.1145/3511176.3511199

[2] Chen Huihong, Liu Shiming, Hu Yaomin Face recognition technology analysis and system architecture [J]. New industrialization, 2017, 7(002): 26–32. https://doi.org/10.19335/j.cnki.2095-6649.2017.02.005

[3] Zhuang Xianlin. A review of face recognition methods[J]. Science and Technology Innovation and Application, 2022, 12(02):130–132. https://doi.org/10.19981/j.CN23-1581/G3.2022.02.035

[4] Song Chaofeng, Yang Jian, song Wenai, et al Improved convolutional neural network architecture [J]. Computer engineering and design, 2019, 40(3): 6. https://doi.org/10.16208/j.issn1000-7024.2019.03.037

[5] Wang Zhaozheng. Cloud storage technology innovation direction outlook[J]. Technology Information, 2017, 15(30):20–21. https://doi.org/10.16661/j.cnki.1672-3791.2017.30.020

[6] Tian Yanling, Zhang Weitong, Zhang Chishi, Lu Gang, Wu Xiaojun. A review of image scene classification techniques[J]. Journal of Electronics, 2019, 47(04):915–926. https://doi.org/10.3969/j.issn.0372-2112.2019.04.020

[7] Deng Wenwen, Sun Chengming, Qin Peiliang Research on the collection method of cloud storage massive data [J]. Modern electronic technology, 2018, 41(14): 4. https://doi.org/10.16652/j.issn.1004?373x.2018.14.003

[8] Li Xudong, Ye Mao, Li Tao. A review of research on target detection based on convolutional neural networks[J]. Computer Application Research, 2017, 34(10):2881–2886+2891. https://doi.org/10.3969/j.issn.1001-3695.2017.10.001

[9] Guo Ruowei. Analysis of network storage technology based on cloud-oriented data centre [J]. Digital Communication World, 2021(02): 98–99+146. https://doi.org/10.3969/J.ISSN.1672-7274.2021.02.041

[10] Zheng le Communication and effect of mobile phone photography in the context of new media [J]. New media research, 2017, 3(6): 2. https://doi.org/10.16604/j.cnki.issn2096-0360.2017.06.009

[11] Ma Linmei, Tang Xiao, Wang Yinhe A new method of image recognition based on dual network model [J]. Software guide, 2020, 19(9): 5. https://doi.org/10.11907/rjdk.201119

[12] Jiang Jie, Xiong Chang town A fine-grained classification algorithm for data enhancement and multi model integration [J]. Journal of graphics, 2018, 39(2): 7. https://doi.org/10.11996/JG.j.2095-302X.2018020244

[13] Lv Haoyuan, Yu Lu, Zhou Xingyu, et al. A review of semi supervised deep learning image classification methods [J]. Computer science and exploration, 2021, 15(6): 11. https://doi.org/10.3778/j.issn.1673-9418.2011020

[14] Ji Changqing, Gao Zhiyong, Qin Jing, Wang Zumin. A review of image classification algorithms based on convolutional neural networks [J/OL]. Computer Applications: 1–6 [2022-01-23]. https://doi.org/10.11772/j.issn.1001-9081.2021071273

[15] Fu Xuyang Analysis of network data storage system under cloud computing technology [J]. Computer programming skills and maintenance, 2019(7): 3. https://doi.org/10.16184/j.cnki.comprg.2019.07.031

[16] Su Yonggang, Gao Jiaqin A fast image retrieval algorithm based on Transfer Learning [J]. Modern computer, 2020. https://doi.org/10.3969/j.issn.1007-1423.2020.34.016

[17] Wang Xiaoyu, Han Changlin, Hu Xinhao Dense connected network face recognition algorithm based on weighted feature fusion [J]. Computer science and exploration, 2019, 13(7): 11. https://doi.org/10.3778/j.issn.1673-9418.1812016

[18] Zhao WQ, Yang PP. Two-way feature fusion combined with attention mechanism for target detection[J]. Journal of Intelligent Systems. 2021(06). https://doi.org/10.11992/tis.202012029

[19] Zhang Shan, Liang Qian, Yang Qian, Hao Peng, Lai Jincui. Application Research of Enterprise Cloud Storage Platform[J]. China Management Informatization, 2022, 25(02):85–87. https://doi.org/10.3969/j.issn.1673-0194.2022.02.028

[20] Luo Mingzhu, Xiao yewei Research on multiscale face detection based on full convolution neural network [J]. Computer engineering and application, 2019, 55(5): 6. https://doi.org/10.3778/j.issn.1002-8331.1805-0034

[21] Li Huiyuan Requirement analysis and modeling of photo album management system based on UML [J]. Information technology and informatization, 2021. https://doi.org/10.3969/j.issn.1672-9528.2021.08.045

Biographies

Zhendong Feng majored in information management and information system in the School of Mathematics and Computer Science of Guangdong Ocean University. He is the core technician of the university's scientific and technological innovation team, mainly engaged in algorithm research in the field of artificial intelligence and information security. He participated in many scientific and technological innovation and entrepreneurship projects with the team, and led the team to participate in discipline competitions for many times. It has rich project practice experience and outstanding scientific research achievements in the direction of computer vision.

Wei Liu majored in software engineering at Guangdong Ocean University. He undertook a national level project of the Innovation and Entrepreneurship Training Plan for Chinese College Students, which is mainly dedicated to building a prediction system for sudden infectious diseases in 2022. He has some research experience in the direction of sudden infectious disease prediction, and has undertaken relevant solutions. He is currently committed to using artificial intelligence to help people make better decisions and efficient solutions in the face of diseases.

Yinghuai Yu Master of Guizhou University Associate Professor of Guangdong Ocean University, member of Zhanjiang Computer Society, with profound research in the field of computer vision. He has published several academic papers in related fields and has served as a technical consultant for several projects. Has excellent experience in project guidance.