深圳大学考试答题纸

(以论文、报告等形式考核专用)  
二○ 一九 ～二○ 二○ 学年度第 2 学期

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 课程编号 |  | | 课序号 | |  | | 课程名称 | |  | | 主讲教师 | 刘凌 | | 评分 |  |
| 学 号 |  | | | 姓名 | |  | | 专业年级 | |  | | | | | |
|  | | | | | | | | | | | | | | | |
| 教师评语： | | | | | | | | | | | | | | | |
| 题目： | | Single-Label Image Classification | | | | | | | | | | |  | | |

**Single-Label Image Classification**

**Abstract**

Image classification is a hot-research area in recent decades. It is so powerful in most domains that it attracts many people to do researches. The most basic image categorization technique is single-label image classification. It can be divided into three different levels. Each level has its own degree of difficulty. One of the most powerful tools to classify images is Neural Network (NN). This paper will mostly focus on Convolutional Neural Network (CNN) and introduce some related famous image databases. Some indicators like Top-5 and Confusion Matrix to evaluate the model are also illustrated in this paper.

**Keywords**

image classification; neural network; indicators; deep learning

1. **Introduction**

Due to the fast development of Internet and wide spread of media information, etc., massive images are generated every day. Lots of them can be classified into different categories, such as the sunset, pets and various food. Thus, scientists thought whether we can ask machines to classify images automatically to help us make better judgements. For example, with the help of cancer cells image classification, doctors can better determine whether the patient suffer the breast cancer. In factories where mechanical parts are produced, workers can better judge whether the parts are qualified with the assistance of parts image categorization. If the categorization technique marks each image with a predictive tag, it is called single-label image classification. This technique can be applied in three different levels: species level, subclass fine-grained level and instance level. With the level goes further, the difficulty of image categorization increases dramatically.

Lots of researches have been done on image classification. Many categorization methods have been proposed and studied, such as K-Nearest Neighbor (KNN), Convolutional Neural Network (CNN), Support Vector Machine(SVM), Back Propagation (BP) and Transfer Learning. Among them, CNN is the most popular and deeply-studied technique in recent years. With the further researches of neural networks and the development of deep learning, training and testing of classification models have become easier and more efficient to be implemented.

Once CNN finishes training the model, some indicators should be given to evaluate the model. Most people will think of accuracy at once. Nevertheless, accuracy is not the only indicator. There are some more complicated indicators to evaluate the classification model.

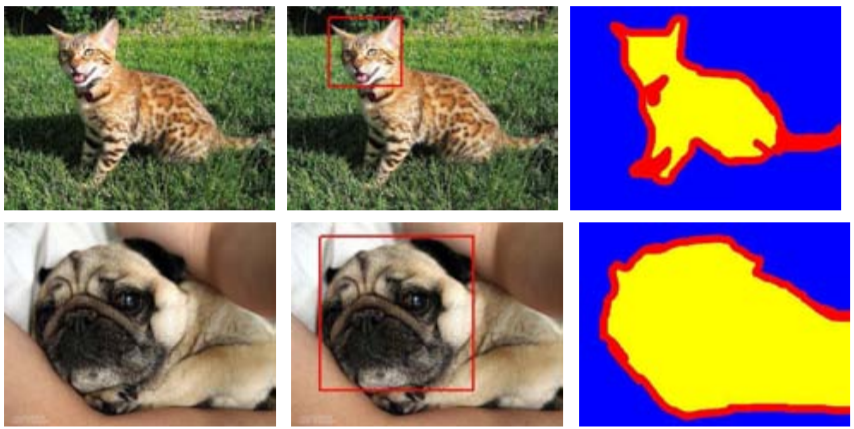
The rest of this survey paper is organized as follows: Section 2 gives a description of three different levels of image classification; Section 3 illustrates two famous image databases and some kinds of CNN used to train classification models; Section 4 introduces some important indicators to evaluate the categorization models; Section 5 draws conclusions of this survey paper.

1. **Three Levels of Image Classification**

**2.1****. Species Level**

This classification works on species level, such as dogs and cat, cars and planes, insects and plants. The aim of this categorization technique is to classify different species. Due to the fact that there exist obvious differences among different species, the classification will be easier to be executed than the next two levels.

Take dogs and cats for example. The classification model will extract stable and distinctive components of the body, such as the face and the main appearance. Fig. 1 shows the components extraction of cats and dogs. With these useful components information, the model can almost obtain an encouraging result to classify dogs and cats [1].

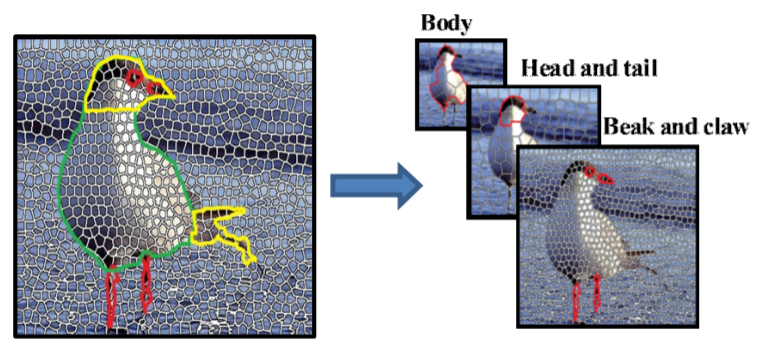


**Fig. 1: Components extraction of cats and dogs**

**2.2. Subclass Fine-grained Level**

This classification is aimed at classifying different subspecies in the same species. This will be harder than classifying across species level.

Take classification on birds images for example [2]. Due to issues like similar shapes and appearances, identification of different subspecies of birds is not an easy work. Therefore, in order to achieve a well-behaved model, one of the most important work is feature extraction. To extract valid features, fine-grained recognition is needed. Through fine-grained classification, the model can discriminate similar birds with subtle differences. Fig. 2 shows the effect of a fine-grained-processed image of a bird. The left side shows a single super-pixel mosaic describing the bird and its parts. The right side shows researchers’ improvements on feature extraction based on the fine-grained image. As we can see, it describes different sized parts in different layers, which will make the model more efficient and accurate.



**Figure 2: Fine-grained-processed image of a bird**

**2.3. Instance Level**

This kind of classification is more challenging! It aims to classify different instances in the same subspecies, such as different flags of each country, different persons like Bob and Alice.

Face recognition is the best example on this level. Advanced feature extraction algorithms can also be applied. The key steps of face recognition are face detection and feature extraction. These features can be local features such as lines and fiducial points, or facial features such as eyes, nose and mouth. But the largely unsolved problem is how to achieve a satisfying model with face images acquired in an outdoor environment with changes in illumination and pose. In other words, current systems are still far away from the capacity of the human perception system [3].



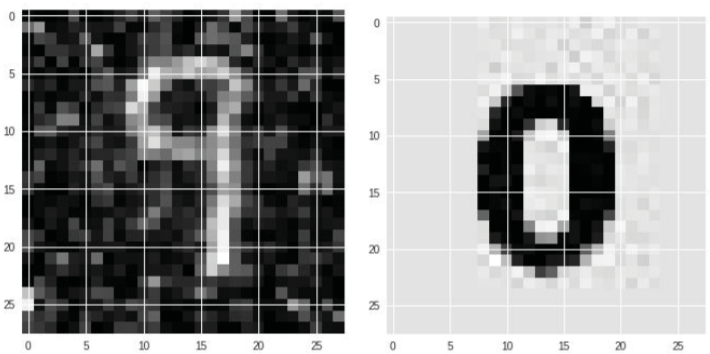
**Fig. 3: Face detection**

**3. Image Databases and different CNNs**

With the development of database technique, some representative image databases had appeared. How to categorize these huge image data has aroused many researchers’ interest. To solve these problems, many excellent CNNs have been designed and applied.

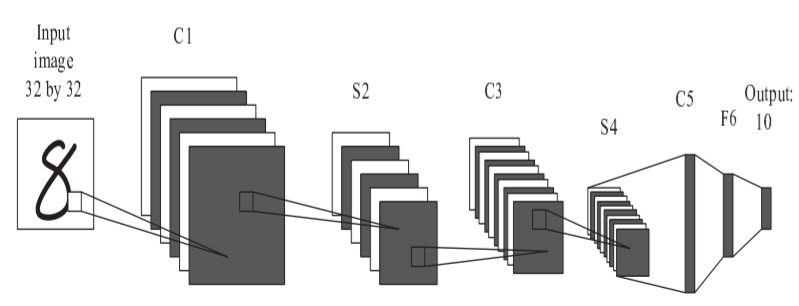
**3.1. MNIST and LeNet5**

MNIST is a classic and basic dataset which has a training set of 60,000 images and a test set of 10,000 images. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image [4]. Many models use it as the first database to train handwritten text images categorization.



**Fig. 4: Sample image data in MNIST**

LeNet5, which originates from LeNet proposed by LeCun in 1994, is a common model of CNN. It was first aimed at classifying handwritten digit images based on MNIST. This network can select feature containing the most distinguishable characteristics among different classes [5]. The structure of LeNet5 is shown in Fig. 5.



**Fig. 5: Structure of LeNet5(C: convolutional layer, S: Subsampling layer)**

**3.2. ImageNet and AlexNet**

ImageNet, which is also a huge database, consists of over 15 million labeled high-resolution images in roughly 22,000 categories. The images were collected from the web and labeled by human labelers using Amazon’s Mechanical Turk crowd-sourcing tools [6]. The most famous competition in Computer Vision is ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which uses a subset of ImageNet with roughly 1000 images in each of 1000 categories.

AlexNet is the most studied CNN. It was designed by Alex Krizhevsky and won the championship in ILSVRC in 2012. It is a deep convolutional network, which contains eight learned layers, five convolutional layers and three fully-connected layers. It tradeoffs between speed and accuracy. With these years’ development, AlexNet can yield 4096-dimensional feature vector for each image and thus achieve higher accuracy [7].

**3.3. Other classification models**

After AlexNet was proposed, the research in deep learning developed faster. More excellent training network appeared, including Clarifai, zfnet, VGGNet, GoogLeNet, ResNet, DenseNet. etc. Among them, GoogLeNet is hot and famous.

GoogLeNet was developed recently based on deep convolutional neural network. It was first proposed by Christian Szegedy and won the championship in ILSVRC in 2014. It is an efficient deep neural network architecture, which has a new level of organization called “Inception Module”. Compared with VGG network, which won the second place in ILSVRC in the same year, the parameters of GoogLeNet are less and its accuracy is better [8].

**4. Indicators for Classification Model**

When the model finishes training, it will begin testing. This section gives some evaluated indicators for image classification models.

**4.1. Top-5 Error Rate**

Now suppose an image in test set is selected. The model receives the image input and gives its predictions after processing. In most cases, the model will give five predictions, which are arranged from high to low in order of the probability. Fig.6 shows an example of the Top-5 predictions to the color of the chair.



**Figure 6: Top-5 predictions to the color of the chair**

The Top-5 error rate is the percentage of the time that the classifier did not include the correct class among its top 5 guesses [9]. Similarly, we can give the definition of Top-1, Top-2, etc. easily.

**4.2. Accuracy Rate**

Accuracy indicates the number of correct predictions made divided by the total number of predictions made [9]. For example, if the classifier totally makes 100 predictions and 95 of them are correct, then the accuracy is:

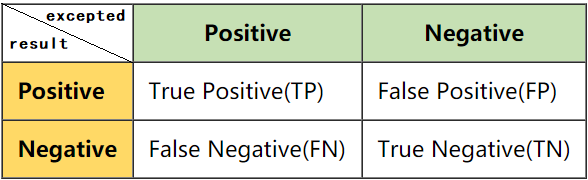
(1)

This indicator has its shortcoming. If the proportion of the sample dataset is imbalance, the accuracy may be inaccurate in reality.

**4.3. Confusion Matrix**

Confusion matrix is a table to present the prediction results of a classifier. In a binary classification, 0 represents negative and 1 represents positive. Table. 1 shows a confusion matrix in a binary classification.

**Table 1: Sample Confusion Matrix in a binary classification**



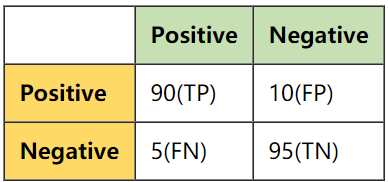
The top side indicates the expected values and the left side indicates the result values. If the result value corresponds to the expected value, then it is ‘true’. Otherwise, it is ‘false’.

**4.4. Precision rate**

Precision is quite similar with accuracy literally, but they are different indeed. Precision is the number of TP(True Positive) divided by the total number of TP and FP(False Positive). It reflects the precision in positive samples rather than all samples. Table. 2 shows a sample confusion matrix. The Precision can be calculated as formula (2).

(2)

**Table 2: Sample Confusion Matrix**



**4.5. Recall rate**

Recall rate is the number of TP divided by the number of TP and FN. Consider the matrix above, the recall rate is,

(3)

It can be a measure of a classifier completeness.

**4.6. F1-score**

F1-score conveys the balance between the precision rate and recall rate. It is calculated as formula (4).

(4)

**5. Conclusions**

Single-label image classification is the kernel topic in this paper. Firstly, we have learned three different levels in single-label image categorization, which are Species Level, Subclass Fine-grained Level and Instance Level. Secondly, we learn about some famous image databases and some excellent CNNs such as LeNet5, AlexNet and GoogLeNet. Lastly, we learn about some indicators to evaluate the trained models. Although image classification technique has developed fast in the past, current image classification algorithms have reached limits and they are still not perfect. Future work should be done to overcome the existing problems, such as the imbalance proportion of image dataset, big variance of the database on fine-grained level, etc.

**6. Acknowledgements**

This survey paper has been written under the instructions of Professor Ling Liu. Most of the reference papers were got from ACM/IEEE by the Digital Library of Shenzhen University.

**References**

1. O. M. Parkhi, A. Vedaldi, A. Zisserman and C. V. Jawahar, "Cats and dogs," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 3498-3505.
2. Luming Zhang, Yi Yang, and Roger Zimmermann. 2015. Fine-Grained Image Categorization by Localizing TinyObject Parts from Unannotated Images. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (ICMR ’15). Association for Computing Machinery, New York, NY, USA, 107–114.
3. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. 2003. Face recognition: A literature survey. ACM Comput. Surv. 35, 4 (December 2003), 399–458.
4. Yann LeCun, Corinna Cortes, Christopher J.C. Burges. (1998). THE MNIST DATABASE of handwritten digits. [Online]. Available: http://yann.lecun.com/exdb/mnist/
5. N. Yu, P. Jiao and Y. Zheng, "Handwritten digits recognition base on improved LeNet5," The 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, 2015, pp. 4871-4875.
6. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90.
7. A. A. Almisreb, N. Jamil and N. M. Din, "Utilizing AlexNet Deep Transfer Learning for Ear Recognition," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), Kota Kinabalu, 2018, pp. 1-5.
8. Ashutosh Singla, Lin Yuan, and Touradj Ebrahimi. 2016. Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model. In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management (MADiMa ’16). Association for Computing Machinery, New York, NY, USA, 3–11.
9. Jason Brownlee. (March 21, 2014). Classification Accuracy is Not Enough: More Performance Measures You Can Use. [Online].Avaliable:https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/