The Development of Image Classification Models Based on Computer Vision

Lingtao Zeng *

Department of Mathematics and Statistics, the University of Manchester, England, Manchester, UK

* Corresponding Author Email: lingtao.zeng@student.manchester.ac.uk

Abstract. Image classification is a fundamental problem in computer vision, which deserves much attention in the past decade. To provide a comprehensive recall for the image classification task based on the deep learning algorithms, this paper first provides a brief history of computer vision and convolutional neural networks. Then, the current state of research and the direction of development of deep learning-based convolutional neural network models for image classification are examined. In addition, the basic model structure, convolution feature extraction, and pooling operations of standard and convolutional neural networks are also introduced. This paper summarizes the development of convolutional neural network models in recent years. In summary, based on the review of the development of image classification algorithms at home and abroad, the current mainstream image classification algorithms and frontier progress are summarized and analyzed, and the existing problems and future development directions of image classification are summarized and prospected. It can be concluded that deep learning-based methods has a great potential in this case.

Keywords: Image Classification; Computer Vision; Deep Learning; Convolutional Neural Network.

1. Introduction

Computer vision is a field of Artificial Intelligence (AI), which refers to making it possible for computers and systems to act or make suggestions using information taken from photos, videos, and other visual inputs. Computer vision empowers computers to discover, observe, and understand. A popular area of study in the field of computer vision is image classification. It aims to determine the category of the image based on the given input image. Currently, image classification algorithms can be observed everywhere e.g. trash classification and medical image classification.

Image preprocessing, feature extraction, and classifier construction make up the main steps in image classification. Image filtering, such as median [1], mean [2], and Gaussian filters [3], as well as image normalization, are all included in the process of preparing images. Its main goals are to purge some unnecessary details from the image, retain as much of the crucial information as possible while condensing the data, and increase the precision of feature extraction. The most important step in the image classification process is feature extraction, which involves transforming the input image in accordance with predetermined rules to produce a new image with specific features. New features typically have low dimension, low redundancy, the benefits of low noise, and are structured, which reduces the need for classifier complexity and boosts model performance. Finally, a classifier is trained to classify the extracted features, so as to realize the image classification.

With the advent of the era of deep learning, the algorithm called neural network achieved very promising performance in many studies. Thereinto, a typical type of the neural network called convolutional neural network was fully employed for dealing with the tasks related to the computer vision [4]. For instance, Alakwaa et al. proposed a 3D-CNN for lung cancer detection and the results demonstrated the effectiveness of it [5]. Qiu et al. proposed a CNN-based Siamese Network for pose matching and achieved the satisfactory results [6].

The aim of this paper is to review the recent progresses of the deep learning-based methods for image classification. Firstly, some datasets that are widely used in various areas are summarized. Then, the basic structure of the convolutional neural network is introduced in detail. Some common convolutional neural network structures e.g. LeNet, AlexNet and GoogleNet are also shown subsequently.

2. Datasets

The datasets used in this field can be divided into the many categories. As the example, several typical datasets in the medical and agricultural domain are summarized briefly. Some sample images can be found in Figure 1 and Figure 2.

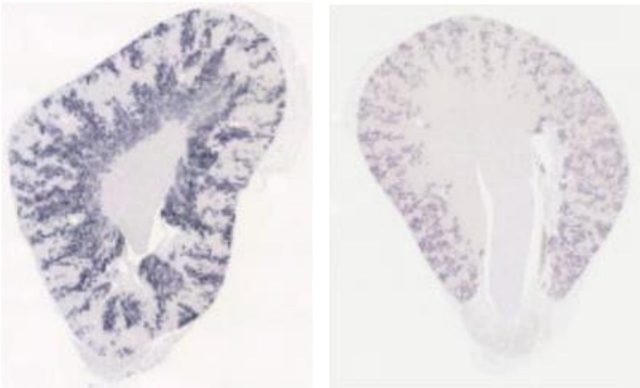

Figure 1. The sample medical images collected from CLEF [7].

Figure 2. The sample agricultural images collected from Agriculture Fruits 360 [8].

2.1. ImageCLEF-2016

This dataset is categorized into 30 groups, including 12 groups for drawings like figures, tables, and flow charts, and 18 groups for medical diagnostic images like CT, MRI, and PET images.

2.2. Patch Camelyon benchmark

327,680 color images (96 × 96 pixels each) from lymph node segment histopathologic scans make up this image classification dataset. Each picture has a binary label that says whether it contains metastatic tissue or not. PCam, which is trainable on a single GPU, larger than CIFAR10, and smaller than imagenet, offers a new benchmark for machine learning models.

2.3. Agriculture TensorFlow_sun397

This dataset contains76,128 training images, 10,875 validation images of 397 categories and 21,750 test images.

2.4. Agriculture Fruits 360

There are 90,483 photos of various fruits and vegetables in this dataset. The test set contains 22,688 photos from 131 distinct classes, whereas the training set contains 67,692 images (one fruit or vegetable per image).

3. The Structure of CNN

Convolutional Neural Networks are a typical type of Neural Networks, which has been widely used in the last decade. It is mainly consisted of three components called convolutional layer, pooling layer and fully connected layer [9, 10]. The convolutional layer will extract features from the input images or feature maps, then activation function will be added to provide the non-linear processing. Pooling layer will be used to reduce the parameters of the CNN to speed up the training process. Fully connected layer will summarize previous features and output the final prediction results. More details of these three components are described below.

3.1. Convolutional Layer

The back-propagation method optimizes the parameters of each convolutional unit that makes up a convolutional layer. The first convolutional layer can be used to extract some low-level features like lines, edges, and so on. The deeper network can fulfill the convolutional operation's objective, which is to extract multiple types of characteristics from the supplied image, by obtaining more complex information through integrative.

3.2. Pooling Layer

Another crucial component of the Convolutional Neural Network is pooling. The core of pooling is essentially sampling; nevertheless, the pooling reduces the dimension and compresses to, in some ways, speed up the calculation. The most well-known technique is "max pooling," which divides the supplied image into numerous rectangular portions and outputs the greatest value from each region. Figure 2 depicts the most common pooling layer, which has a stride of 2 and a window size of 2*2 in two dimensions. The layer can extract a maximum of four numbers from each 2*2 fragment of the image.

In addition to the max pooling method, the pooling layer can also use additional functions like "average pooling" and "L-2 norm pooling," among others. In the past, average pooling was widely used, but because the max pooling approach performs better, its use has diminished.

3.3. Fully Connected Layer

The fully connected layer, which follows the convolutional and pooling layers, is where the algorithm does high-level inference. The fully connected layer maps the feature expression vector obtained from earlier levels of feature extraction to the subsequent layer or the final layer.

4. Typical CNNs

4.1. LeNet

LeNet-5 was born in 1998 and is one of the earliest convolutional neural networks completed by Yann LeCun, which has promoted the development of the field of deep learning [11]. At that time, none of the GPU to help train model, even the CPU speed is slow, so LeNet5 through clever design, using convolution, parameters, sharing and pooling operations such as feature extraction, avoid a lot of computation cost, and then use full connection neural network to classify recognition, the starting point of this network is also recently a lot of neural network architecture, has brought a lot of inspiration to the field.

4.2. AlexNet

As the champion network of ILSVRC2012, AlexNet has a similar network structure as LeNet, but with deeper layers and more parameters (about 240M) [12].

AlexNet is an 8-layer convolutional neural network with the first 5 layers as convolutional layers and the last 3 layers as fully connected layers, where softmax is used for classification in the last layer. In this model, the Rectified Linear Units (ReLU) is used to replace the traditional Sigmoid and Tanh functions as the nonlinear activation functions of neurons, and a Dropout method (random inactivation) is proposed to alleviate the overfitting problem and reduce the computational complexity of the training model.

4.3. GoogleNet

Different sizes of convolution kernels mean different sizes of receptive fields. VGGnet has proved that different receptive fields help to improve classification accuracy. GoogLeNet of Google Research goes a step further by convolution kernels of different sizes in the same layer, and such a structure is called Inception [13]. As shown in figure 5, 1×1, 3× 3,5 ×5 convolution kernels of different sizes are inserted in the same layer, and finally stitched to fuse feature images of different scales with each other. The main purpose of adding a 1×1 convolution kernel before 3×3, 5×5 is to reduce the dimension, which can reduce the number of parameters. GoogLeNet structure: The overall network structure is shown in Figure 9, with a total of 22 layers of network: initially composed of 3 layers of ordinary convolutions; Next, it consists of three groups of subnetworks, the first group of subnetworks contains two Inception modules, the second group contains five Inception modules, and the third group contains two Inception modules. Then the mean pooling layer and the fully connected layer are connected.

5. Conclusions

This paper summarizes the overview of CNN and introduces the datasets in various domains as well as some classic models based on CNN e.g. LeNet, AlexNet and GoogleNet. However, there are still certain issues with CNN that need to be fixed. 1) A major problem with using CNNs in new domains is that choosing the right hyperparameters requires a lot of prior practical experience and is expensive due to their internal dependencies. It is currently unclear how to manage these parameters more effectively. 2) Immediate attention is needed for application-specific problems such the network model's reliance on strong hardware, vast amounts of data, and the training models' dynamic adaptability. There are still numerous development opportunities to be explored and realized in the fields of CNN and deep learning in the future.

References

[1] Huang, T., Yang, G. J. T. G. Y., & Tang, G. A fast two-dimensional median filtering algorithm [J]. IEEE transactions on acoustics, speech, and signal processing, 1979, 27(1), 13-18.

[2] Pan, J. J., Tang, Y. Y., & Pan, B. C. The algorithm of fast mean filtering [C]. In 2007 International Conference on Wavelet Analysis and Pattern Recognition (Vol. 1, pp. 244-248), 2007, IEEE.

[3] Kovesi, P. Fast almost-gaussian filtering [C]. In 2010 International Conference on Digital Image Computing: Techniques and Applications (pp. 121-125), 2010 IEEE.

[4] Albawi, S., Mohammed, T. A., & Al-Zawi, S. Understanding of a convolutional neural network [C]. In 2017 international conference on engineering and technology (ICET) (pp. 1-6), 2017, Ieee.

[5] Alakwaa, W., Nassef, M., & Badr, A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN) [J]. International Journal of Advanced Computer Science and Applications, 2017, 8(8).

[6] Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training [J]. Biomedical Signal Processing and Control, 72, 2020, 103323.

[7] Genome Biology https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2441470/figure/F6/, 2008.

[8] Medium https://medium.com/@arnabbhakta956/fruits-and-vegetables-classification-with-fruits-360-dataset-using-deep-learning-36610f9c2122, 2022.

[9] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. Convolutional neural networks: an overview and application in radiology [J]. Insights into imaging, 9(4), 611-629, 2018.

[10] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. A survey of convolutional neural networks: analysis, applications, and prospects [J]. IEEE transactions on neural networks and learning systems, 2021.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition [C]. Proceedings of the IEEE, 86(11), 2278-2324, 1998.

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 60(6), 84-90, 2017.

[13] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. Going deeper with convolutions [C]. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9), 2015.