2022 International Conference on Big Data, Information and Computer Network (BDICN) | 978-1-6654-8476-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/BDICN55575.2022.00134

Image Classification based on CNN: Models and Modules

Haoran Tang*
Chongqing Biacademy
Chongqing, China
*aiden.tang@biacademy.cn

***Abstract*—With the recent development of deep learning techniques, deep learning methods are widely used in image classification tasks, especially for those based on convolutional neural networks (CNN). In this paper, a general overview on the image classification tasks will be presented. Besides, the differences and contributions to essential progress in the image classification tasks of the deep learning models including LeNet, AlexNet, Inception, VggNet and ResNet are introduced. This paper will also explain in detail, how different units in these CNN models, other than the convolutional layer, including pooling, activation, and dropout functionalize to support better results for these models. These results offer a guideline for deeply understanding the utility of CNN units.**

Keywords-component;Image clasiification; CNN; Deep Learning

I. Introduction

Basically, image classification is a type of tasks that tries to understand the image and classify the image into a specific category. While this sounds an extremely easy task for human beings to manually decide the class of different images, image classification in the data science field means automatically processing a huge number of images in parallel and provides the classification results simultaneously. Meanwhile, while judging a common viewed image like animals can be simple for humans' eyes, the image classification task could be significantly difficult for most people if such image needs specific knowledge to classify, e.g., the x-ray photo. In this case, the image classification task for a machine means to completely comprehend images as data, and classify the image based on information that can be barely understood by human beings, like the relative position of pixels of the image. As the recent development of deep learning techniques has processing to the improvement of the image classification task, this report will provide a comprehensive review on the image classification task and examine how different CNN based deep learning models can achieve the desirable results for image classification. Then, this paper will introduce how different units in the CNN models besides the convolutional layer itself, play their unique roles in processing the image data.

II.OVERVIEW OF THE IMAGE CLASSIFICATION TASK

A.Hisotry of Image Classification

Although the image classification task has only become explosively popular among different various academic fields and industry in recent decade, the image classification tasks itself is actually not a strange term to computer scientists. Image classification is counted as one subfield of the computer visions, which can be traced back to 1960s when computer scientists tried to conduct research on the visual architecture of visual shapes of various images. Since then, there were frequent research on visual perception and recognition, until entering early 21st century, when there was a shift of the focus to object detection and image classification tasks in computer vision after a few research and experiment on visual classification with CNN based models in 1990s. One of the most successful work as the start of the image classification tasks was LeNet, developed by Yann LeCun, who was viewed as the pioneer to apply CNN and backpropagation based algorithms to establish an effective neural architecture for the image classification task. Following LeNet, more neural networks extended from LeNet were created to achieve better results on the image classification task. Represented by AlexNet, CNN based deep learning models further determined their fame and raised the popularity of the image classification research. Meanwhile, with the development of the computation power of CPUs and GPUs, deeper neural networks and more variation of CNN based deep learning models were created. Scientists and practitioners have made great efforts in developing advanced classification approaches and techniques for improving classification accuracy to meet the challenge of more complexity and wider landscape in this field. A variety of CNN based deep learning models has surged for the image classification task including Inception, Vggnet and Reset, which have explored into the advanced complexity of CNN, aiming to provide stronger predictability of the model.

B.Process of Image Classification

While significant focus and reports have been developed on the deep learning modeling, the image classification task involves a serious of traceable tasks instead of merely building a complex CNN based model. Similar to most deep learning tasks, the first step of the image classification task is data processing and feature engineering. This step means to prepare a clean set training and test data with clearly defined features and targets as the input for the deep learning model. Unlike other machine learning tasks, the image classification tasks are special as the image data is mostly more complex considering its three dimensional property, but more simple features as features are mostly just pixels of the image. That is why the feature engineering of the modern image classification will also including the possible choice of data augmentation and join of other side information of the images. The next step is the core to build and train the deep and customized deep learning model suitable for the specific image classification task. The modern deep learning models for the image classification task are mostly

978-1-6654-8476-3/22/$31.00 ©2022 IEEE
DOI 10.1109/BDICN55575.2022.00134

CNN with their strength in weight sharing and parameter reduction. The training process could involve the finetune process to identify the best parameter choice for the final deep learning model. Lastly, the model will be evaluated on certain metrics, mostly AUC or accuracy, for the image classification task, to decide which model to be implemented for this specific task.

III.CNN BASED DEEP LEARNING MODELS

In this section, a few representative CNN based deep learning models that have been provide with well-performed results on the image classification task will be introduced.

A.LeNet

LeNet is one of the earliest approaches combined the CNN and propagation to help the deep learning model to learn the feature representation of the image and predict the class of the image developed by Yann LeCun [1]. The main idea behind the application of CNN into the deep learning model is to extract the low-dimensional feature representation from the image. Figure 1 shows the architecture of LeNet. Through the learning of feature map, all the weights learning with backpropagation can be synthesized and reduce the capacity of the machine by the weight sharing technique (LeCun et al., 1998)[1]. LeNet basically starts the era of applying CNN based deep learning models in the image classification task, and it has achieved competitive predicative results in digit classification. The architecture is show in Figure 1.

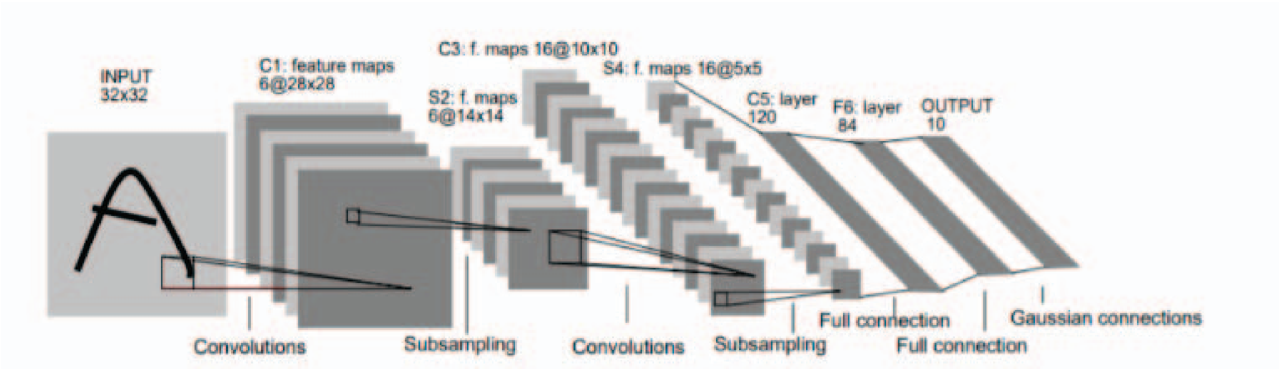

Figure 1. The Architecture of LeNet

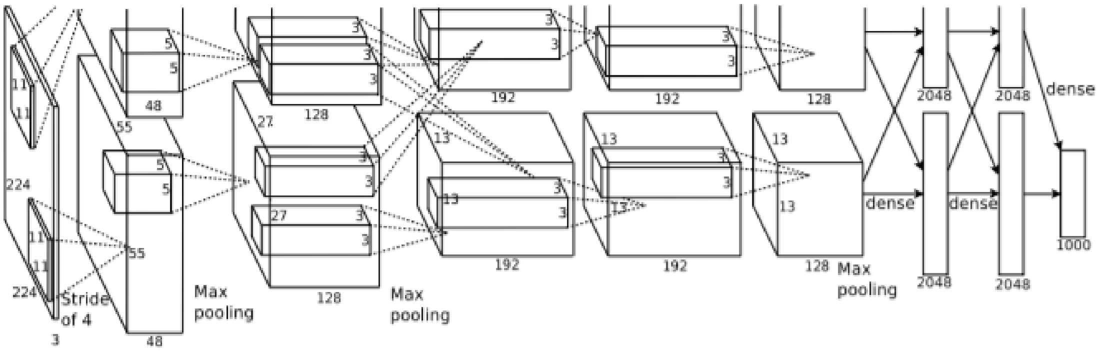

Figure 2. The Architecture of AlexNet

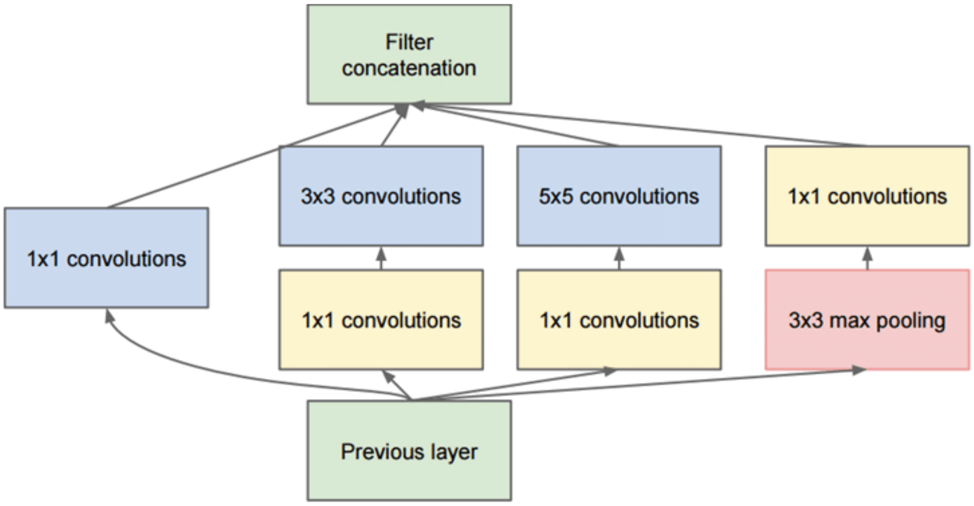

Figure 3. GoogLeNet Inception Architecture [3]

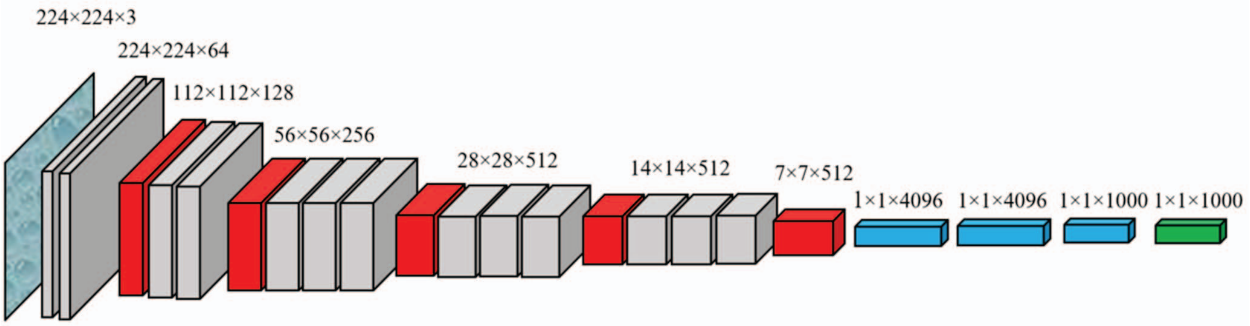

Figure 4. The Architecture of VggNet

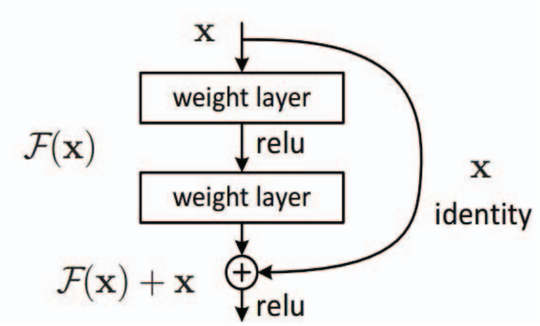
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 5. Residual Learning Block

B.AlexNet

AlexNet is basically an extension of LeNet, which has used deeper CNN based structure within its framework. Figure 2 has shown the architecture of AlexNet. Other than being a deeper model, there were three main tricks used by AlexNet. AlexNet uses ReLU instead of the traditional tanh as its activation unit for the nonlinearity to accelerate the backpropagation learning process. The second trick is data augmentation to provide additional feature by transforming the original image data, and the third trick is to apply dropout, i.e., to randomly set the hidden neuron as 0 with a 0.5 probability. The latter two techniques have mainly generated some noise into the data and the training to reduce the overfitting of the complex model architecture [2].

C.Inception

The inception technique, created by scientist at Google, is based on the idea to combine more varied information in the convolutional layer. Basically, in the convolutional layer, instead of only relying on ono feature map, inception applies multiple filters and concatenates the results together as the next input to extract multi-level feature at each step. To reduce the parameter count, Inception models have factorized smaller size of convolutions to reduce to cost of adjacent tiles.[4] In addition, inception has also taken the advantage of the auxiliary classifier to further regularize the deep learning model.

D.VggNet

The idea behind VggNet is simple, which is deeper layers of the CNN based deep learning model can yield better results. VggNet is based on the fact that stacking more layers can help the deep learning model identify and extract more information with the receptive fields from the image. One specific implementation detail is its initialization of the model. It first trains a deep learning model with a relatively shallow network configuration and initializes the deeper architecture partly with the result of the shallow network. This is to circumvent the problem that bad initialization may stall the learning due to the gradient instability in deep nets. Figure 4 shows the architecture of VggNet [5].

E.ResNet

A residual neural network (ResNet) is an artificial neural network (ANN) of a kind that builds on constructs known from pyramidal cells in the cerebral cortex. Residual neural networks do this by utilizing skip connections, or shortcuts to jump over some layers. Typical ResNet models are implemented with double- or triple- layer skips that contain nonlinearities (ReLU) and batch normalization in between [6]。 An additional weight matrix may be used to learn the skip weights; these models are known as HighwayNets [7]. Models with several parallel skips are referred to as DenseNets [8].

There are two main reasons to add skip connections: to avoid the problem of vanishing gradients, or to mitigate the Degradation (accuracy saturation) problem; where adding more layers to a suitably deep model leads to higher training error [1].

Compared to the efforts of going deeper to achieve advanced complexity for the stronger predictability, ResNet aims to avoid overfitting brought by the extremely deep architecture that may stop the model from developing better performance. ResNet develops the concept of the residual learning that connects the deeper part with the idendity mapping of the shallow part to prevent the degradation of deeper model [9]. Figure 5 shows a residual learning block. The connection between layers strengthens the learning of the deep learning model on different levels of features to maintain the stability of the deep neural nets.

IV.Deep learning Modules

A.Pooling

Pooling is the process to further aggregate the output from the convolutional layer, with the commonly used techniques as the max pooling and the average pooling. As an intuitive explanation, polling can basically summarize the information extracted from the previous convolutional layer, and further reduce the number of computations need for the feature map. Implicitly, pooling acts as a robust change to maintain the output of the convolutional layer less affected by the location of the pixels of the image.

B.Activation

Activation is a fundamental unit for almost all the deep learning models, not only limited to the CNN based models. As deep learning model can be viewed as the combination of different linear transformations, the activation unit is to add the nonlinearity property for the deep learning model to simulate the complex information within the data. Commonly used activation functions include sigmoid, tanh, and ReLU, where ReLU is the most frequently used one in the image classification task as it can effectively mitigate the gradient vanishing or the gradient exploding problems in the training process.

C.Dropout

Dropout, as a useful trick used in AlexNet, is a method to randomly drop out the hidden nodes of the deep learning model, mostly with zero. Dropout can be extremely useful for the image classification task with significantly deep neural network architecture as it can avoid potential overfitting by inducing additional data noise and regularizing the model by preventing the extremely large weights within the model. Dropout is a common hyperparameter in terms of the probability of dropping out hidden neurons. It is simple, yet effective way to help the deep CNN based model to better generalize during the training.

V.Conclusion

In summary, the image classification method based on CNN model is systematically reviewed. First, the history and definition of the task is introduced. Then, the process of image classification task is discussed. Subsequently, the corresponding contemporarily CNN approaches are demonstrated accordingly since they are the core. In addition, some modules in CNN will also greatly affect the performance of the model. Therefore, the module's function and effectiveness are discussed. Overall, these results shed light on the in-depth understanding of CNN.

References

[1] Lecun, Y. , and L. Bottou . "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11(1998):2278-2324.

[2] Technicolor T , Related S , Technicolor T , et al. ImageNet Classification with Deep Convolutional Neural Networks. NIPS, 2012

[3] Maad Ebrahim., "Performance study of augmentation techniques for HEp2 CNN classification.Apr.2018

[6] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2016). "Deep Residual Learning for Image Recognition" (PDF). Proc. Computer Vision and Pattern Recognition (CVPR), IEEE. Retrieved 2020-04-23

[7] Srivastava, Rupesh Kumar; Greff, Klaus; Schmidhuber, Jürgen. "Highway Networks". May 2011.

[8]Huang, Gao; Liu, Zhuang; Weinberger, Kilian Q.; van der Maaten, Laurens (2016-08-24). "Densely Connected Convolutional Networks"

[4] Szegedy C , Liu W , Jia Y , et al. Going Deeper with Convolutions[J]. IEEE Computer Society, 2014.

[5] Szegedy C , Vanhoucke V , Ioffe S , et al. Rethinking the Inception Architecture for Computer Vision[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:2818-2826.

[9] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).